Revenue Management for Airline Bookings using Regression Analysis

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Technology

by

Vaibhav B. Bhange

to the

DEPARTMENT OF INDUSTRIAL AND MANAGEMENT ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY KANPUR
INDIA

June, 2005

CERTIFICATE

This is to certify that the work contained in the Thesis titled "**Revenue Management for Airline Bookings Using Regression Analysis**" has been carried out by **Vaibhav Bhange** under my supervision and that his work has not been submitted elsewhere for a degree.

June, 2005

Dr. A. K. Mittal,

Professor,

Dept of Industrial and Management Engineering,

Indian Institute of Technology, Kanpur-208016.

ACKOWLEDGEMENT

I would like to express deep gratitude to my supervisor Dr. Ashok K. Mittal for his invaluable guidance and encouragement, which inspired me throughout. He continuously encouraged me through his valuable suggestions.

I would also like to thank Dr. Parthasarathy Ramachandran for his initial guidance for six months in understanding and developing the Simulated system which was at the central part of the study without which it would have been very difficult to complete the present work.

I extend my thanks to the entire IME faculty and staff for providing an atmosphere that helped me in promoting my personal and professional growth and making the learning into wonderful experience.

Last but not the least, I would like to thank my batch mates of IME department and friends of Hall-IV and IITK, who helped me in times of crisis and supported me in the times of need and made my stay at IIT Kanpur a memorable one. I wish them all a great future.

June, 2005 Vaibhav B. Bhange

Abstract

Under the purview of revenue management seat inventory control problem, the current study by means of Simulated System, Experimental Design and Regression analysis proposes two new models to find optimal protection level for higher fare class against lower fare class in a nested booking structure of airline reservations system for a single leg seat optimization problem. The fare distributions for the two classes are considered so that the airline will be able to carry those passengers who have potential to maximize the revenue. It is expected that the new model developed will help improve the revenue management of the airline industry to some extent.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Problem Statement

The air traffic has been increased since the deregulation of US airline industry in 1978. The number of air travelers is increasing day by day. To cope up with this increasing demand and to grab the opportunity, various new small players are entering into this competitive market. The practice of seat inventory management enable airlines to influence their total revenues on a flight-by-flight basis, within a given price structure.

The application of booking limits on the number of seats available at different prices on the same flight in the same coach, allows airlines to increase revenues. Effective seat inventory control by an airline depends on forecasts of future bookings, the revenue values associated with each fare type, and an ability to make systematic tradeoffs between booking requests so as to maximize total flight revenues. The Expected marginal Seat Revenue (EMSR) decision model developed by Peter P. Belobaba (1989) [1] exists that takes into account the uncertainty associated with estimates of future demand as well as the nested structure of booking limits in airline reservations systems. EMSR was the first widely applied optimization-based methodology for controlling the seat inventory.

However the problem with EMSR model is that it considers only the average fares of different classes to get the protection levels for higher fare classes and not their spread [P. P. Belobaba (1989)] [1]. Another method known as Dispersed Fare Rule (DFR) developed by Larry Weatherford (2003) [2] takes advantage of the variability of actual fare values around the mean fare values to generate a significant revenue improvement over EMSR (2–6 per cent). This heuristic uses a threshold mechanism. Same is the case with Leg Bid Price (LBP) heuristic due to Larry Weatherford (2002) [3].

In case where fares have distribution, in EMSR system, a request with higher fare paying low class customer can get rejected, while the seat may go empty because of non-availability of the passengers from higher class. The main aim of this thesis is

to develop a heuristic which will take into consideration this phenomenon and provide a reasonable protection level to extract more revenue from the potential low class customers. This thesis is aimed at developing a method as the extension of EMSR model which takes into account the fare distribution of various fare classes to get the protection levels to optimize revenue gains.

1.2 Scope of the Thesis

This thesis is aimed at improving the traditional model of seat inventory control in revenue management to squeeze out the revenue from the extra seat available taking into consideration fare distributions. A model is developed that is likely to provide airline industry a tool to improve its revenue management.

1.3 Thesis organization

This study is broadly divided into four chapters. The first chapter gives an overview of the need to improve the traditional model. It emphasizes the motivation for improvement. The problem statement is specified along with a brief discussion on the scope of thesis.

Second chapter reviews the literature on revenue management and previous works done in the field of airline seat inventory control. The Expected Marginal Seat Revenue (EMSR) decision model is discussed at length in this chapter. The Dispersed Fare Rule (DFR) and Leg Bid Price (LBP) which takes into account the spread of fare levels in each class are also discussed.

Third chapter introduces the methodology adopted for the study. This chapter also discusses the Design of Experiments and Analysis of Variance methods which were extensively used in the study. Experiments performed and their results are discussed. The two models which take into account fare distribution and demand distribution are developed and compared with the EMSR model.

Fourth chapter concludes the findings along with the scope for future research required in this area.

Chapter 2

Literature Review

The airline revenue management problem has been studied extensively by the academics and their work has had much practical impact. In this chapter we review the literature in this field, in particular on the seat inventory control problem and the decision models associated with it.

2.1 Revenue Management

2.1.1 Background

This section is based on excellent research overview done by McGill and Van Ryzin (1999) [4]. Before the deregulation of the U.S. airline industry in 1978, airlines typically only offered a single fare in the markets (city pairs) they served. This fare was set by the Civil Aeronautics Board (CAB), using a mileage-based formula that guaranteed airlines a healthy profit margin in return for providing air transport to the nation. These fares were increased each year in response to the rising industry cost, leaving airlines with no pressing incentive to operate efficiently [see Sanne Vincent de Boer (2003)] [9].

The CAB also determined which airlines could operate which routes, limiting the number of carriers serving each market. As a result, the only way airlines could compete for passengers was based on service, both in-flight and in terms of the departure times and frequency of their flights. Since higher frequencies make air travel a more convenient option, an airline that adds flights on a particular route can expect a higher share in a larger market, at little or no risk of not recovering the cost. This factor contributes to the problem of overcapacity, which unfortunately is all too common in this industry. Many flights take off with empty seats, known as *spoilage* since the opportunity to sell them has passed.

The bulk of the operating costs in the airline industry is fixed, while the variable costs of carrying an additional passenger are relatively small. In the short run, it therefore makes sense for an airline to sell seats that would otherwise remain empty

at a large discount, higher than the marginal cost that they contribute to recovering the fixed costs. The key issue is that such discounted fares should not result in loss of passengers who would otherwise have paid higher-fare. This would lead to *diversion* of existing high-fare passengers, *spilling* revenue as a result. [(Sanne Vincent de Boer (2003)] [9]

In the early 1970s, some airlines began offering restricted discount *fare products* that mixed discount and higher fare passengers in the same aircraft compartments. For example, BOAC (now British Airways) offered earlybird bookings that charged lower fares to passengers who booked at least twenty-one days in advance of flight departure. This innovation offered the airline the potential of gaining revenue from seats that would otherwise fly empty; however, it presented them with the problem of determining the number of seats that should be *protected* for late booking, full fare passengers. If too few seats were protected, the airline would *spill* full fare passengers; if too many were protected, flights would depart with empty seats. No simple rule, like protecting a fixed percentage of capacity, could be applied across all flights because passenger booking behavior varied widely with relative fares, *itineraries*, season, day of week, time of day, and other factors.

It was evident that effective control of discount seats would require detailed tracking of booking histories, expansion of information system capabilities, and careful research and development of *seat inventory control* rules. Littlewood (1972) [10] of BOAC proposed that discount fare bookings should be accepted as long as their revenue value exceeded the *expected revenue* of future full fare bookings. This simple, two fare, seat inventory control rule (henceforth, *Littlewood's rule*) marked the beginning of what came to be called *yield management* and, later, *revenue management*.

Yield is the revenue per passenger-mile of traffic carried by an airline. An airline's yield will be a function of both the prices it charges for its differentiated service options and the number of seats actually sold at each price. Yield management thus involves two major components- pricing and seat inventory control, Larry Weatherford (1998) [6].

Over the last twenty years, development of revenue management systems has progressed from simple *single leg control*, through *segment control*, and finally to

origin– destination control. Each of these advances has required investment in more sophisticated information systems, but the return on these investments has been excellent. In 1999, most of the world's major air carriers and many smaller airlines have some level of revenue management capability. Other small airlines and international airlines in newly deregulated markets are beginning the development process. The success of airline revenue management was widely reported, and this stimulated development of revenue management systems for other transportation sectors and in other areas of the service sector.

2.1.2 Revenue Management by Domestic Airlines in India

In recent times, Indian airline companies like Indian Airlines, Jet Airways, Air Deccan, SpiceJet, Air Sahara, Kingfisher Airlines, etc. have started offering discounted fares. There is a long list of discounted fares provided by these companies. Some discounted fares are general whereas some are specific to the company. APEX fares, Super Saver fares, Night saver fares, Check fares, etc are some of the examples of the discounted fares offered by Indian companies.

APEX (Advance Purchase Excursion) fares were introduced in India only during 2002. Apex fares are usually available with all the airlines and requires advance purchase of the ticket. While Jet allows you to book 15, 21 and 30 days in advance and Indian Airlines 7, 21 and 28 days, Sahara has gone in for 5, 10, 15, 20 and 30-day booking periods. For Jet, out of the roughly 20,000 passengers that travel everyday on an average, 4,000 travel on apex fares. In case of Indian Airlines roughly 2,000 passengers opt for apex fares out of a total of over 35,000 passengers that fly every day, (websites) [26, 27, 28].

Jet Airways offers 'Check Fares' which are 30 to 40 per cent lower than the regular Economy class fares. Unlike Apex fares, for 'Check Fares' one does not need to book in advance and the reservation and ticketing has to be done simultaneously. Similar is the concept of 'Steal Fare' scheme launched by Air Sahara and 'Select Fares' by Indian Airlines. The fares offered by Air Deccan are 50 per cent cheaper than the fares offered by other airlines on some of the routes. Further fare varies with the time of booking. SpiceJet, the latest private domestic low-cost air carrier, has recently launched the "Red Hot Special Fares" which they claim are the lowest fares of all the times, (websites) [27, 28, 29, 31].

There is a frequent flyer program under which a customer can get a free return ticket on flying certain number of return flights. For this one has to be the member of flying return program and score the mileage points. Jet Airways call it Jet Privilege program and use a new approach called Dynamic Tier Review instead of traditional miles based points. Similar program is called as Power Flying Program by Air Sahara whereas Indian Airlines calls it flying return. The rate at which one gets the free ticket is different for different airlines, (websites) [26, 27, 28].

2.1.3 The Airline Revenue Management Problem

This section is based on the research overview done by McGill and Van Ryzin (1999) [4]. The objective in revenue management is to maximize profits; however, airline short-term costs are largely fixed, and variable costs per passenger are small; thus, in most situations, it is sufficient to seek *booking policies* that maximize revenues. Also, although there is lower risk in accepting a current booking request than in waiting for later possible bookings, booking decisions are repeated millions of times per year; therefore, a risk-neutral approach is justified.

Consider the arrival of a booking request that requires seats in an itinerary—one or more flights departing and arriving at specified times, within a specific *booking class*, at a given fare. The fundamental revenue management decision is whether or not to accept or reject this booking. Durham (1995) [11] reports that a large computer reservations system must handle five thousand such transactions per second at peak times, thus the decision must be reached within milliseconds of the request's arrival. Not surprisingly, no current revenue management system attempts full assessment of each booking request in real time. Instead, precomputed aggregate control limits are set that will close the system for further bookings of specific types while leaving it open for others. The reservations system can quickly determine the open or closed status of a booking category and report back to an agent or customer without actually evaluating the request.

The accept–reject decision can be restated as a question of valuation: What is the expected displacement cost of closing the incremental seats in the requested itinerary? To maximize expected revenues, the request should be satisfied only if the fare value of the requested itinerary equals or exceeds the expected displacement cost (see Talluri and Van Ryzin 1999b)

The apparent simplicity of this valuation problem is deceptive—a complete assessment must allow for all possible future realizations of the reservations process that could be influenced by the availability of any of the seats on any of the legs in the booking. Fully traced, this influence propagates across the entire airline network because a booking can displace potential bookings that will have subsequent impacts of their own. This influence also propagates forward in time because many affected itineraries will terminate later than the booking being considered. Also, a booking will normally have a return component at a later date with its own set of concurrent and downstream effects.

The performance of a given revenue management system depends, in large part, on the frequency and accuracy of updates to control limits and the number of distinct booking classes that can be controlled. The determination of suitable control limits and characterization of their *structural* properties over time has been the principal focus of academic research, whereas the need for practical and implementable approximations to optimal limits has driven much of the practitioner research.

2.1.4 Forecasting

This section is based on the research overview done by McGill and Van Ryzin (1999) [4]. Forecasting is an important component of planning in any enterprise; but it is particularly critical in airline revenue management because of the direct influence forecasts have on the booking limits that determine airline profits. Not surprisingly, publication of approaches to airline forecasting are concurrent with the literature on overbooking because overbooking calculations depend on predictions of ultimate demand, cancellations, and no-shows. Unfortunately, the disaggregate forecasting required for both overbooking and revenue management is extremely difficult. Simply accounting for the effects of price volatility is a significant challenge in itself.

When we talk about *Models for Demand Distributions,* empirical studies have shown that the normal probability distribution gives a good continuous approximation to aggregate airline demand distributions. Given the central limit theorem and the role of the normal distribution as the limiting distribution for both binomial and Poisson distributions, this is not surprising. However, many researchers have pointed out that

the normal distribution becomes increasingly inappropriate at greater levels of disaggregation.

Most of the research works use a *Model for the stochastic arrival process* of individual booking requests to construct distributions of total flight demand. In other work that seeks dynamic booking rules, specification of the arrivals process is an essential starting point. The use of the Poisson process, when appropriate, is useful in dynamic treatments because of the memoryless property of the exponential interarrival distribution; however, both the homogeneous and nonhomogeneous Poisson processes lead to Poisson cumulative arrival distributions. This is problematic for total demand modeling because the *coefficient of variation* of the Poisson distribution is the reciprocal of the square root of the mean.

Researchers like Swan, Lee and McGill have worked a lot in the area of *Uncensoring Demand Data.* Data contained in historical booking records are *censored* by the presence of booking and capacity limits on past demands. Swan (1990) [13] addresses the downward bias of censoring on late booking data and suggests simple statistical remedial measures. An earlier *spill formula* developed by Swan has been used for many years by practitioners to *unconstrain* demand. Lee (1990) [14] presents a detailed stochastic model of passenger arrivals based on a censored Poisson process and develops maximum likelihood methods for estimating the parameters of these models. McGill (1995) [15] develops a multivariate multiple regression methodology for removing the effects of censorship in multiple booking classes, and describes a bootstrapping approach to testing for correlations between fare class demands.

Research has shown that use of regression techniques can improve the performance of revenue management systems when compared to time series analysis or historical averages. Airlines are understandably reluctant to share information about their forecasting methodologies because their revenue management activities are so heavily dependent on accurate forecasting. As far as we know at this time, most disaggregate forecasting systems depend on relatively simple moving average and smoothing techniques augmented with careful analysis of recent booking profiles, as mentioned above. Manual intervention is required on an exception basis for critical markets or to anticipate the impact of changes in prices or other important aspects of market structure. Regression and time series techniques have proven of some use for

forecasts of aggregate demand, but not at the disaggregate level. This mirrors the general state of forecasting methodology in inventory control applications reported widely throughout industry.

2.1.5 Overbooking

This section is based on the research overview done by McGill and Van Ryzin (1999) [4]. Before the ground breaking work by Littlewood (1972) [10] on seat inventory control, almost all quantitative research into airline revenue management focused on optimal overbooking. The issue is to find the right balance between the opportunity cost of empty seats for which demand was turned down (spoilage) and the cost of having to "bump" (deny boarding) a number of passengers when a flight is oversold. The latter consists of the direct compensation offered to these passengers (cash, flight vouchers), but also of lost goodwill, which may lead passengers to choose a different airline in the future. Again, research in this field can be divided into static models and dynamic models. In static models, the airline sets a fixed overbooking level at the beginning of the booking period, which determines up to how many reservations it can have accepted at any given time. The dynamic model focuses on individual booking requests over time, which can now be accepted even if the number of confirmed reservations already exceeds the capacity of the flight.

2.1.6 Pricing

This section is based on the research overview done by McGill and Van Ryzin (1999) [4]. Most part of the literature deals with pricing and price competition at an industry level rather than the operational, revenue management decision level. This literature is nonetheless relevant to strategic and marketing decisions that are important in revenue management.

Dana (1996) [16] shows that a firm that offers high and low prices and then rations the capacity sold at the low price, as is done in yield management practice, is in a unique competitive equilibrium. Borenstein and Rose (1994) [17] provide empirical tests of airline competition and its relationship to the degree of price dispersion observed in fares.

It is now common for airline practitioners to view pricing as part of the revenue management process. The reason for this is clear—the existence of

differential pricing for airline seats is the starting point for revenue management, and price is generally the most important determinant of passenger demand behavior. There is also a natural duality between price and seat allocation decisions. If price is viewed as a variable that can be controlled on a continuous basis, a booking class can be shut down by raising the price sufficiently high. Also, when there are many booking classes available, shutting down a booking class can be viewed as changing the price structure faced by the customer.

Weatherford (1994) [18] presents a formulation of the simultaneous pricing/allocation decision that assumes normally distributed demands, and models mean demand as a linear function of price. Recent work by Gallego (1996) [19] uses a simple deterministic model to examine pricing and market segmentation decisions. His model takes into account both demand diversion and demand recapture. He gives precise conditions to guarantee the optimality of low to high pricing and lower and upper bounds on the optimal revenue.

2.1.7 Seat Inventory Control

This section is based on the research overview done by McGill and Van Ryzin (1999) [4]. The problem of seat inventory control across multiple fare classes has been studied by many researchers since 1972. There is a progression from Littlewood's rule for two fare classes, to *expected marginal seat revenue* (EMSR) control for multiple classes, to *optimal booking limits* for single-leg flights, to segment control and, more recently, to *ODF control.* The single leg and network problems can be reviewed separately.

2.1.7.1 Single-Leg Seat Inventory Control

Most early seat inventory control research required most or all of the following simplifying assumptions:

1) S*equential booking classes*;

2) L*ow-before-high fare* booking *arrival pattern*;

3) Statistical independence of demands between booking classes;

4) No cancellations or no-shows (hence, no overbooking);

5) Single flight leg with no consideration of *network effects*; and,

6) No batch booking.

Littlewood's rule can be viewed as an early expression of the displacement cost rule for two booking classes under all six assumptions. Belobaba (1987a) p[25] extends Littlewood's rule to multiple fare classes and introduces the term EMSR for the general approach. The EMSR method does not produce optimal booking limits except in the two-fare case; however, it is particularly easy to implement. A later refinement of EMSR, called EMSRb, apparently produces better approximations to optimal booking limits and has been widely implemented. Curry (1990) [20] proposes an approximation for the network problem, a relaxation of assumption 5. Research has been done to characterize optimal booking limits with a set of probability conditions that are the same as the EMSR conditions for the first two fare classes but include joint probabilities that are lacking from the EMSR method for additional fare classes.

2.1.7.2 Segment and Origin–Destination Control

Since the 1980s, network effects in revenue management have become increasingly significant because the expansion of hub-and-spoke networks has dramatically increased the number of passenger itineraries that involve connections to different flights. It has been recognized for some time that revenue management should account for these network effects but that this cannot be accomplished effectively with single-leg control. Progress in this area has been impeded not by the lack of approaches to network inventory control so much as by the limitations of older reservations systems. Even if optimal solutions existed for the many hundreds of itineraries that traversed a single leg, those solutions had to be mapped into a much smaller number of *controllable booking classes*. This situation is changing—the most advanced reservations systems are now capable of incorporating network information, and the emergence of *seamless availability* will allow for much finer control of seat availability. We review, below, four approaches that have been taken to network revenue management.

Mathematical Programming Formulations: Early approaches in the passenger O-D problem were considered as minimum cost network flow formulations assumed deterministic demands so that they can focus on network effects rather than stochastic elements. Wollmer (1986c) [21] proposes a linear programming (LP) network formulation that allows for stochastic demand by incorporating expected marginal

seat values as coefficients in the objective function. Each ODF generates a set of zero–one decision variables for each flight, with a corresponding set of monotonically decreasing objective function coefficients determined by the marginal expected values. Wollmer shows that this formulation can be converted to a minimum cost network formulation for greater efficiency in solution; however, the size of the problem for a typical airline network is extremely large.

One of the drawbacks of these mathematical programming formulations is that they produce non-nested allocations. The formulations have seen some use for planning purposes but have not been implemented for day-by-day seat inventory control. Their main potential seems to be as components of the bid-price approaches.

Curry (1990) [20] describes a combined mathematical programming/marginal analysis formulation for the O–D problem that uses piecewise linear approximations to the revenue function in a linear program that obtains distinct bucket allocations for different O–Ds. The different fare classes for each O–D are then nested, and each O–D nest is separately optimized for single-leg, nested booking limits. This approach has been implemented in some revenue management systems.

Segment Control: The earliest implementations of partial O–D control were at the flight *segment* level. These implementations allow for the revenue value of a multi-leg itinerary as long as the itinerary does not involve connections between different flights. The motivation for this partial solution to the ODF control problem was the feasibility of exploiting *segment closed indicators* that were available in the reservation control system. The methods available to determine seat/segment allocation rules are similar to those available for the broader O–D control.

Virtual Nesting: The first systems that addressed the broader O–D control problem were developed to accommodate the limited number of controllable booking classes that the CRS (Computer Reservations System) provided. Techniques for clustering ODFs (Origin Destination Fare) into single-leg booking classes to achieve an approximation to network control were developed. Such methods assign ODFs to booking classes on the basis of some measure of their total value to the airline instead of just their fare class.

A variety of options are available for the clustering process (often called *indexing*), including assignment by total value, assignment by estimated leg value

(prorated, for example, by leg distance), and assignment by estimated net value after allowance for displacement effects (dual prices from a deterministic network LP).

Bid-Price Methods: Smith and Penn (1988) [22], Simpson (1989) [23], and Williamson (1992) [24] incorporate information from LP/network models into the detailed accept–deny decision process of seat inventory control. They use dual prices from a deterministic LP model to establish marginal values for incremental seats on different legs in an airline network. Typically, expected demands replace random demands as constraints in the LP formulations. The dual prices are summed across legs in a passenger itinerary to establish an approximate displacement cost, or *bid-price*, for that itinerary. A booking request for a passenger itinerary is rejected if the bid price exceeds the fare for the itinerary, and is accepted otherwise. Thus, this approach attempts to directly incorporate the estimated displacement cost as a cutoff value for acceptable fares. Despite potential theoretical drawbacks to the bid-price approach, it has a very convincing advantage—it replaces multiple booking limits and complex nesting schemes with a single bid-price value for each flight leg and a simple rule for rejecting or accepting itinerary requests.

2.2 Expected Marginal Seat Revenue Model

Determining the booking limit on each fare class that will maximize total revenues for a future scheduled flight departure is a dynamic process. The static problem is to establish these fare class limits at the start of the booking process, taking into account the uncertainty associated with expected bookings by fare class, to the extent possible. The dynamic problem is to revise these initial limits on the basis of the additional information provided by actual bookings as departure day approaches. The Expected Marginal Seat Revenue (EMSR) model recommends fare class booking limits, taking into account the probabilistic nature of future demand for a flight.

2.2.1 EMSR

This section is based on the research done by P. P. Belobaba (1989) [1]. It is assumed that the lowest fare classes tend to book up first, well before the majority of requests for the highest fare class are received due to the restriction of advance purchase on these lowest fare classes. The seat inventory control problem therefore is

to determine how many seats not to sell in the lowest fare classes and to retain for possible sale in higher fare classes closer to departure day. A decision model therefore must find the protection levels for higher fare classes which can be converted into booking limits on lower fare classes. In a nested fare class reservations system, each booking limit is the maximum number of seats that may be sold to a fare class (including all lower fare classes with their own, smaller booking limits.) The booking limit on the highest fare class is, thus, the total capacity of the shared cabin. The protection level for the highest fare class is the difference between its booking limit and the booking limit of the next lowest class.

So the assumptions taken into consideration for EMSR model are:

1. There is no relationship between demand levels for different fare classes.
2. A customer denied a flight/ fare class request represents a booking loss to the airline.
3. There are no booking cancellations or passenger no-shows occur.
4. All the revenue classes have a single revenue value (e.g. class A customer always pays $300, class B customer $250, etc)
5. The classes always fall into nice hierarchical order i.e. A>B>C>D etc.
6. All the lower-valued price classes book before the higher-valued price classes.

In addition, there is one implied assumptions like the demand estimates for each price class are accurate.

There exists uncertainty about the ultimate number of requests that an airline will receive for seats on a future flight and, more specifically, for the different fare classes offered on that flight. The total demand for a particular flight, on average, fluctuates systematically in cycles described by day of the week and season of the year. There also will be stochastic variation in demand around the expected values, among similar flights sampled consistently over a homogenous period of time. This stochastic demand for a future flight departure can be represented by a probability density function. Past analyses generally have assumed a normal distribution of total demand for a flight, with means and variances that depend on the market being studied and on the nature of traffic.

We define $p_i(r_i)$ to be the probability density function for the total number of requests for reservations, r_i, received by the airline for seats in fare class i by the close of the booking process for a scheduled flight leg departure.

The number of seats allocated to a particular fare class, S_i, might not exceed the number of actual requests for that fare class, resulting in rejected demand, or spill. Thus, we can define a cumulative probability that all requests for a fare class will be accepted as a continuous function of S_i:

$$P_i(S_i) = P[r_i \le S_i] = \int_0^{S_i} pi(ri)dri$$

Conversely,

$$P[r_i > S_i] = \int_{S_i}^{\infty} p_i(r_i)dr_i$$

$$= 1 - P_i(S_i) = \overline{P}_i(S_i)$$

The probability of receiving more than S_i requests for fare class i, or the probability of spill occurring, is therefore $\overline{P}_i(S_i)$

We define $EMSR_i$ to be the expected marginal seat revenue for class i when the number of seats available to that class is increased by one. The expected marginal seat revenue of the S_ith seat in fare class i, $EMSR_i(S_i)$, is simply the average fare level in that class multiplied by the probability of selling S_i or more seats:

$$EMSR_i(S_i) = f_i.\overline{P}_i(S_i)$$

Note that $EMSR_i(S_i)$ depends directly on $\overline{P}_i(S_i)$, the probability that the S_ith seat made available to class i will be sold.

Consider a single-leg flight for which bookings will be accepted in two nested fare classes, 1 and 2, having average fare levels f_1 and f_2 respectively. In order to maximize total expected flight revenues, the reservations process should give priority to class 1 passengers. Class 1 will have the total available capacity of the shared

cabin, C, as its booking limit, BL_1. The seats protected from class 2 and available exclusively to class 1 will be denoted S_2^1. The optimal protection level S_2^1 for class 1 is the value of S_2^1 that satisfies the condition

$$EMSR_1(S_2^1) = f_2$$

Graphically, the optimal value of S_2^1 is the point at which the $EMSR_1(S_1)$ curve intersects f_2, as shown in figure 1. The optimal booking limit on class 2 is BL_2, the difference between the capacity of the shared cabin, C, and the optimal protection level, S_2^1.

This solution will maximize expected revenues in cases where the booking limit is set at the start of the reservations process (static seat inventory control). A class 2 request will be rejected only when BL_2 is reached, at which point, the expected revenue for all remaining seats will be greater than the class 2 average fare. If class 2 requests never reach BL_2, the unsold seats will be available for unexpectedly high class 1 demand. In any event, the expected revenue per seat from class 1 requests in excess of S_2^1 is below f_2, in the absence of additional information on actual bookings for the future flight being managed.

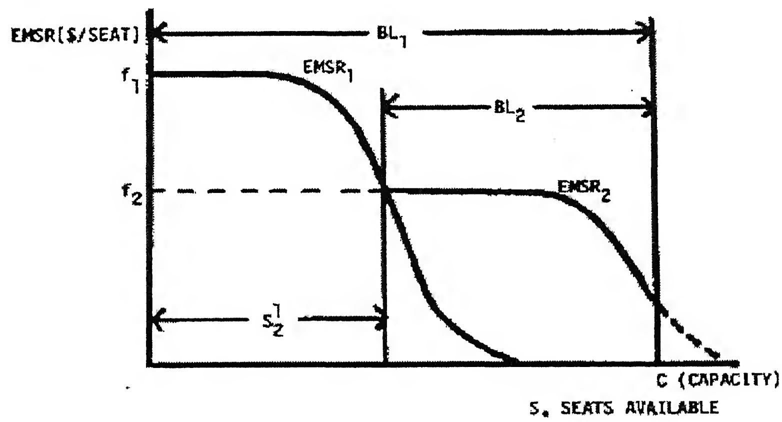


Figure 2.1: Maximizing expected revenues for the 2-class example

Extending this approach to multiple fare classes on a single flight leg simply requires that more comparisons of expected marginal revenues be made among the

relevant classes. In general case of k fare classes offered on a flight leg, the optimal values of S_j^i must satisfy

$$EMSR_i(S_j^i) = f_i, i < j, j = 1,...,k.$$

The total number of comparisons required for k nested fare classes is given by

$$\frac{k(k-1)}{2}.$$

These protection levels, in turn, determine the booking limits on each fare class j:

$$BL_j = C - \sum_{i<j} S_j^i .$$

That is, all seats with an expected marginal revenue greater than f_i should be held back from sale to class j. Otherwise, any request for a class j seat may be accepted. It is possible that one or more values of BL_j derived from these equations might be negative, in which case, class j should not be offered at all if expected revenues are to be maximized. In such a case

$$BL_j = \max[0, C - \sum_{i<j} S_j^i].$$

The incremental number of seats protected for class j is the nested protection level for class j, denoted NP_j. The nested protection level for class j thus is given by

$$NP_j = BL_j - BL_{j+1} .$$

In the EMSR framework, the lowest fare class does not have a protection level per se, but rather a booking limit equal to the number of seats that remain after all upper classes have been protected. With a capacity of C seats and k fare classes, then, the values of NP_j must satisfy

$$C = \sum_{j<k} NP_j + BL_k$$

The EMSR protection levels S_j^i, nested protection levels NP_j, and booking limits BL_j are shown graphically for a three-class example in figure 2.2.

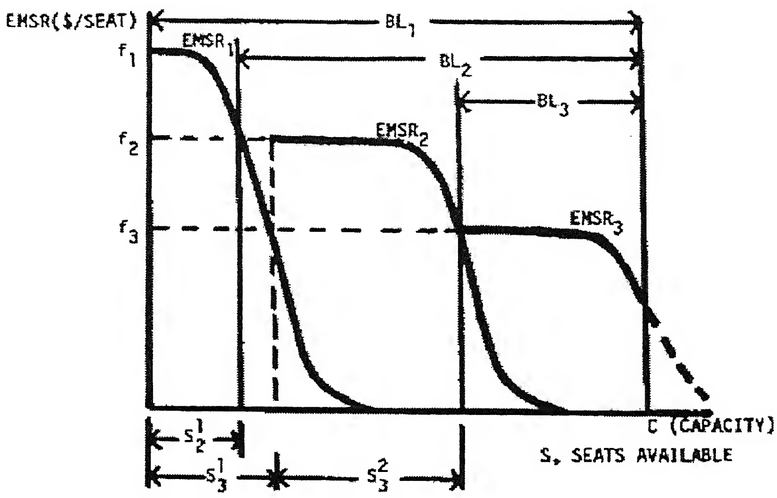


Figure 2.2: EMSR solution for the nested 3-class example.

EMSR booking limits can be generated for any number of nested fare classes with no change to the model's basic structure. It is important to recognize that not all reservations systems are nested in this way, or they might not be nested at all. Nesting is preferable in seat inventory management because there is no difference in the physical seats or the on-board service being sold to different fare classes.

2.2.2 Dynamic Application of the EMSR model

This section is based on the research done by P. P. Belobaba (1989) [1]. The EMSR decision framework can be applied to a seat inventory control context in which booking limits may be revised on a regular basis a s the light departure day nears. In such a situation, additional information is available in the form of actual bookings already accepted for the future light. Because an actual booking in any fare class will translate into a revenue passenger occupying a seat, incorporating actual bookings into the EMSR decision framework can reduce the uncertainty associated with the estimates of expected demand used as input.

In the static problem, the EMSR model requires an estimate of the total expected requests by class. In the dynamic case, estimates of future requests at various times before departure are required to calculate optimal protection levels for the unbooked seats still available for flight. Dynamic application of the EMSR framework involves repetitive use of the static model but with revised input data. The

objective is to determine the optimal fare class limits for the time period optimality of the booking decisions already made.

We define r_i^t to be the number of requests made for class i between days t and 0 before departure meaning:

$$r_i^t \leq r_i$$

The probability density of requests from day t onwards $p_i(r_i^t)$, and the probability of receiving S or more requests fro class i in the time remaining to departure is $\overline{P_i^t}$ (S).

On any day t prior to flight departure, the inputs required by the EMSR model are the average fare or revenue levels for each fare class, f_i, which may or may not remain constant over the booking period, and the estimates of $\overline{P_i^t}$ (S) for all relevant values of S, derived from the $p_i(r_i^t)$ densities. The optimal seat protection level for class 1 relative to class 2 for the period remaining before departure is S_2^1 (t), such that

$$EMSR_1^t[S_2^1(t)] = f_1.\overline{P_1^t}(S_2^1) = f_2$$

This protection level for day t can be used ot find the revised optimal booking limit on class 2, as

$$BL_2(t) = C - b_1^t - S_2^1(t)$$

Where b_1^t is the number of bookings already accepted in class 1 up to day t before departure. The maximum number of seats still available is C - b_1^t and S_2^1 (t) of these seats are protected for class 1.

For more than two fare classes, comparisons between the EMSR(S) values of all upper classes relative to the average fare levels of lower classes are required, a before. These comparisons involve the demand densities of future requests from the current day t. Each comparison of a higher fare class i with a lower fare class j generates an optimal value of S_j^i (t) that satisfies

$$EMSR_i\,[S_j^i(t)] = f_i.\overline{P_i^t}(S_j^i) = f_j$$

The revised booking limits for day t take actual bookings into account:

$$BL_j(t) = C - \sum_{i<j} S_j^i(t) - \sum_{i<j} b_i^t$$

The nested protection levels, $NP_j(t)$, for successively lower fare classes are derived as in the initial case, but with actual bookings included:

$$NP_j(t) = BL_j(t) - BL_{j+1}(t)$$

$$= \sum_{i<j} S_{j+1}^i(t) - \sum_{i<j} S_j^i(t) + b_j^t$$

As before, $BL_j(t)$ is constrained to be greater than or equal to zero. It also is constrained in this case to be no lower than the actual number of bookings already accepted in class j and all lower fare classes are generally received earlier in the booking process than full-fare requests, it is possible that the revised $BL_j(t)$ derived from the EMSR calculations will be lower than the number of bookings already on hand in classes j and lower. With no possibility of cancellations or no-shows assumed, the revised booking limit on class j then becomes

$$BL_j(t) = \max[\, C - \sum_{i<j} S_j^i(t) - \sum_{i<j} b_i^t, \sum_{k \geq j} b_k^t, 0]$$

The EMSR framework thus can be sued to determine optimal protection levels and recommended booking limits in a nested multiple fare class reservations system for a single, future flight leg. Initial booking limits may be derived on the basis of estimates of total expected requests for a future flight, before the reservations process begins. These limits may then be revised dynamically during the booking process, taking into account both actual bookings and estimates of future requests by fare class.

2.2.3 EMSRb Model

This section is based on the research done by P. P. Belobaba (1992) [5]. The basic idea of the EMSRb method is to heuristically break down the multi-fare problem into a sequence of two subproblems. To calculate the EMSRb booking limit corresponding to class $i+1$, the higher classes 1 to i are combined in an artificial class 1,i with demand $r_{1,i}$ and fare $f_{1,i}$.

Under EMSRb, joint protection levels π_i are calculated for all higher classes relative to a given lower class in the following manner:

For class 1: find the value of π_1 that makes $EMSR(\pi_1) = f_1.\overline{P}_1(\pi_1) = f_2$

Where $\overline{P}_i(S_i)$ = probability that $r_i > S_i$, S_i no. of seats made available to class *i*.

the EMSRb expression is same as EMSR for class 1 where $\pi_1 = S_1$

Once π_1 is found, the booking limit for class 2 (BL_2) is set = C- π_1

For class 2: how many seats to protect jointly for classes 1 and 2 from class 3?

The following calculations are necessary:

$$\overline{r}_{1,2} = \overline{r}_1 + \overline{r}_2$$

$$\hat{\sigma}_{1,2} = \sqrt{(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)}$$

$$f_{1,2} = \frac{f_1.\overline{r}_1 + f_2.\overline{r}_2}{\overline{r}_{1,2}}$$

$$\overline{P}_{1,2}(S) = probability[r_1 + r_2 \geq S]$$

we then find the value of π_2 that makes $EMSR_{1,2} = f_{1,2}.\overline{P}_{1,2}(\pi_2) = f_3$

For general class n:

The following calculations are necessary:

$$\overline{r}_{1,n} = \sum_{i=1}^{n} \overline{r}_i$$

$$\hat{\sigma}_{1,n} = \sqrt{\sum_{i=1}^{n} \hat{\sigma}_i^2}$$

$$f_{1,n} = \frac{\sum_{i=1}^{n} f_i \overline{r}_i}{\overline{r}_{1,n}}$$

find the value of π_n that makes $EMSR_{1,n} = f_{1,n}.\overline{P}_{1,n}(\pi_n) = f_{n+1}$

Once π_n is found, set $BL_{n+1} = C - \pi_n$

2.3 Previous works on dispersed fares

This section is based on the research done by Larry Weatherford (2002) [3] and (2003) [2]. All of the previous works done in the area of Airline seat inventory Management have extended the single-leg inventory control optimization problem to relax some of the assumptions made initially in EMSR model with respect to demand independence and order of booking arrivals, all these models depend on two primary inputs, namely forecasts of demand and revenue inputs. With respect to revenue inputs, all the models assume that the revenue value of each booking class is known and in proper hierarchical order with single point value.

Typical fare data for airlines industry as shows that

1. the prices in each fare class follow a probability distribution, rather than the typical assumption of a single fare value, and
2. the average fare values across price classes do not always follow the proper hierarchical order. ($F_k > F_{k+1} > F_{k+2}\ldots$)

Larry Weatherford (2002) [3], made use of the dispersed fares to improve the revenue gain over traditional EMSRb model. He first developed a new heuristic decision rule, called Leg Bid Price (LBP), which is based on Bid Prices. Bid price is often thought of in the airline literature for what it could contribute in a network environment (e.g origin- destination) but it was shown that LBP can outperform EMSRb in a single leg environment. Leg Bid price approach is totally different from EMSRb approach as the former replaces multiple booking limits and complex nesting schemes with a single bid-price value for each flight leg and a simple rule for rejecting or accepting itinerary requests.

In Leg Bid Price a threshold mechanism (i.e. accept the booking if revenue offered by a customer is greater than or equal to the threshold; reject otherwise) is used which controls the bookings. The LBP is the EMR of the last seat allocated that fills the plane's capacity.

While examining LBP, he made use of the data from a major European cruise line with six booking classes A, B, C, D, E, and F. The average fares for each class were not strictly in hierarchical order. Also there was a lot of fare dispersion as shown in figure.

Major assumptions were (1) demand for each booking class is separate and independent of other class demands; (2) demand for each class is stochastic and can be represented by a probability distribution; and (3) the proportion of demand expected to arrive within a given booking period is known.

Shown in figure 2.4 is the example how fares are dispersed and in some cases fare inversion do occur (as in B and C fare classes)

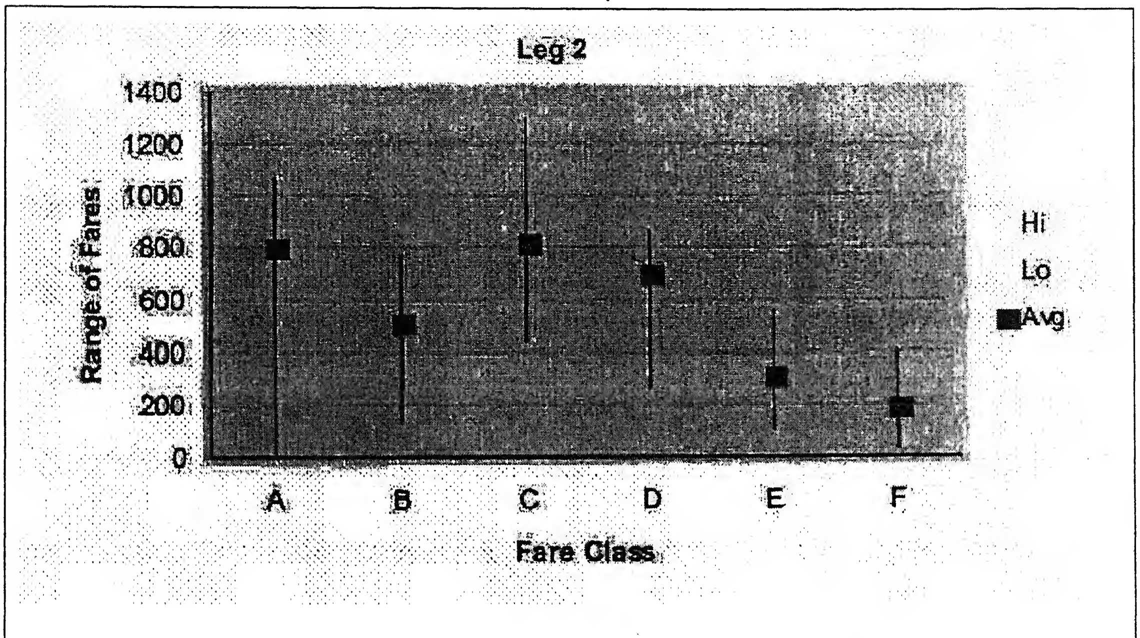


Figure 2.3: Box Plots of fare distributions.

The results of his study showed that the LBP approach improved the revenues to a significant degree; approximately 15 percent. It suggests that increased variance around the mean price values used as inputs to the unit optimization models and greater amounts of fare inversion do have a discernible impact on the revenue generated. [3]

In his second paper on dispersed fares, Larry Weatherford (2003) [2] introduced a new heuristic decision rule called Dispersed Fare Rule (DFR) which showed 2-6 percent revenue improvement over EMSRb.

Basically, the DFR that is calculated to control bookings is the same as the EMSRb method, but with an additional test which uses a threshold mechanism. The way in which the threshold value is set is based on the forecast of demand

relative to capacity and is adjusted such that it spills all demand in excess of capacity.

The study was conducted on two type of legs viz; business leg and leisure leg. The three different methods (i.e. probability distributions) in which the fares might be dispersed within the fare class were considered namely; normal, uniform and skewed. The fares were overlapping, which is quite typical for real companies.

Shown below is the example of Normal distribution of fares.

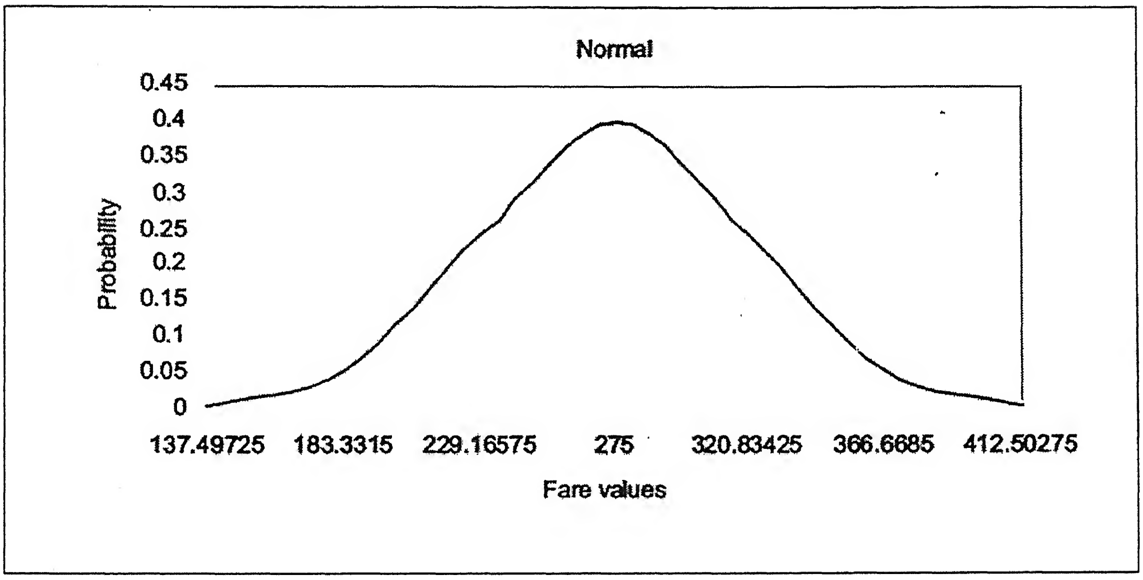


Figure 2.4: Normal distribution of fares in Y class.

The results showed that with greater fare overlap, the potential revenue improvement from the new decision rule was greater.

Chapter 3

Methodology

3.1 Simulation System to Find Optimal Protection Levels

A simulation system for calculating revenue and fare class protection level is designed. The simulation code is written in Java programming language. In this code an input .txt file provides the required input to the program and the output is collected in the excel file. The system is designed with an objective function to compute protection level as per the EMSR heuristic. It can also be used to calculate average revenue for a given fare structure and demand distributions from a specified protection level. The schematic view of the system is as shown in the figure. 3.1

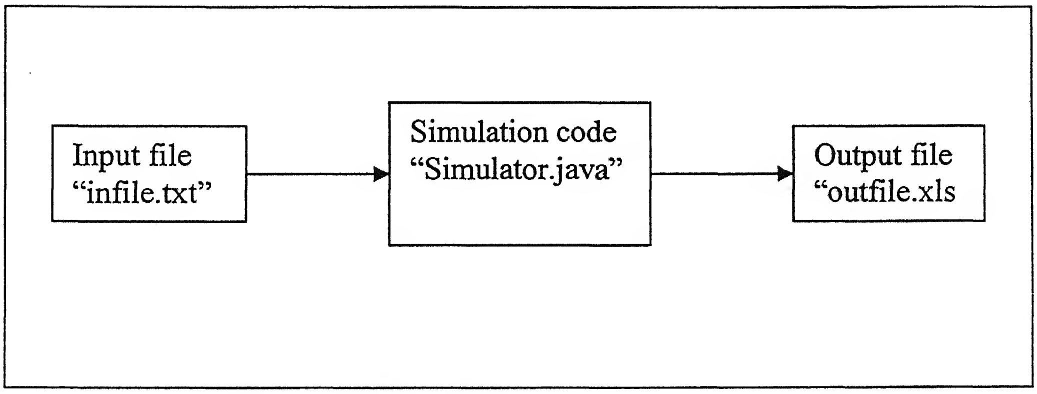


Figure 3.1: Schematic view of the Simulation

3.1.1 Input File “infile.txt”

The input file consists of “n” rows where each row is the unit problem supplied to the program. The unit problem is the values of the parameters required to simulate the reservations booking process. The parameter values provide the demand distribution for two classes and their fare distribution. Notations are as follows:

μ_{d_1} : demand for class 1

μ_{d_2} : demand for class 2

σ_{d_1} : standard deviation of demand for class 1

σ_{d_2} : standard deviation of demand for class 2

μ_{f_1} : mean fare for class 1

μ_{f_2} : mean fare for class 2

σ_{f_1} : standard deviation of fare for class 1

σ_{f_2} : standard deviation of fare for class 2

Following is an illustration of the input file:

	μ_{d_1}	μ_{d_2}	σ_{d_1}	σ_{d_2}	μ_{f_1}	μ_{f_2}	σ_{f_1}	σ_{f_2}
problem1	10	70	5	23	1.1	1	0.33	0.3
problem2	30	85	15	30	1.2	1	0.36	0.3
problem3	24	68	10	20	1.3	1	0.39	0.3

3.1.2 Simulation program "Simulator.java"

3.1.2.1 Computation of Protection Level Using EMSR

This program reads the input file. Each time it reads one line of input file which corresponds to one problem. The parameter values are then extracted from the line read from the file.

The parameter values are used to calculate the protection level for class 1 from the EMSR model. By convention, class 1 customers are considered as high fare paying whereas class 2 customers are low fare paying. EMSR model requires the demand distribution of the two classes and their mean fares. Schema of this module is shown in figure. 3.2

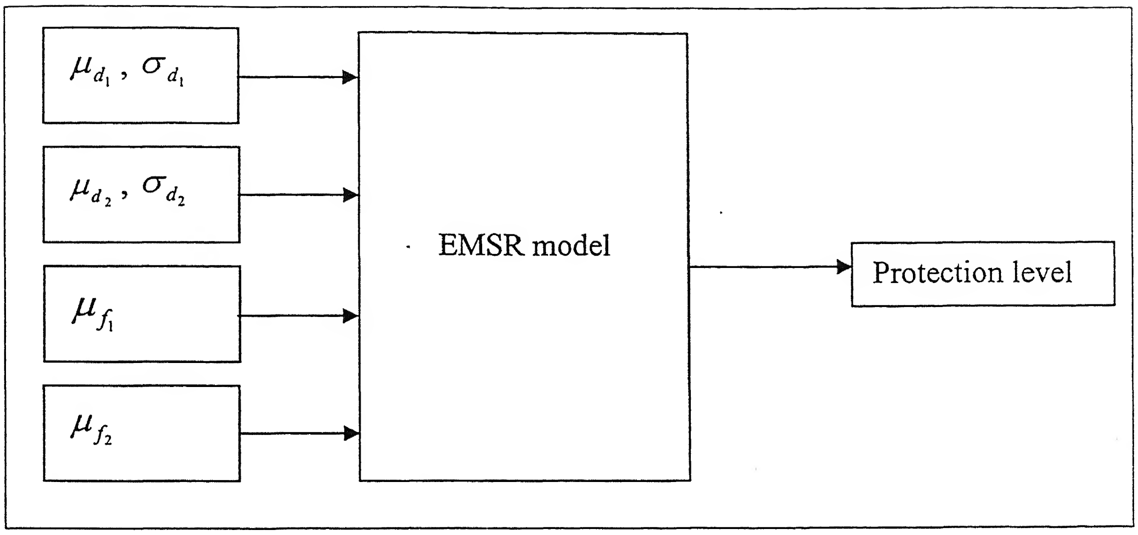


Figure 3.2: Schema of EMSR model

3.1.2.2 Seat Allocation for an Arriving Passenger

Once the protection level is obtained, the next step is to generate the random demand from each type of customer (class1 or 'Y' class and class2 or 'M' class). The demand is generated randomly using distribution provided in the input file i.e. N~(μ_{d_1} , σ_{d_1}) and N~(μ_{d_2} , σ_{d_2}). From this we compute total number of requests which is combined demand for class1 customers and for class 2 customers. Each request consists of following three attributes.

1. Customer type
2. Customer arrival time before the departure time of the flight. and
3. Fare value

- Customer type takes the value either 'Y' or 'M'
- Customer arrival is assumed to be uniform for a specified demand hence the arrival time is computed from uniform distribution
- Fare is computed from the fare distributions, N~(μ_{f_1} , σ_{f_1}) and

  N~(μ_{f_2} , σ_{f_2}) provided by the input file.

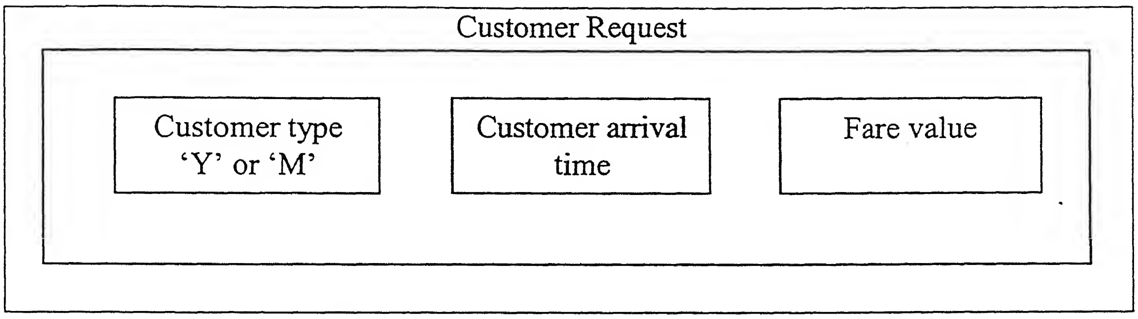


Figure 3.3: Customer Request and its attributes

The simulator computes these three elements of the customer requests as follows:

First, all the class 1 customers are assigned their type i.e. 'Y' and a random value of arrival time. After this is done class 2 customers are assigned their type i.e. 'M' and a random value of arrival time.

All the requests from two classes are combined and then sorted in descending order of arrival time. This signifies that the top most request has been made first. The fare value is then assigned to each request according to its type. The fare values are assigned from the fare distribution. The process is as shown in the figure. 3.4

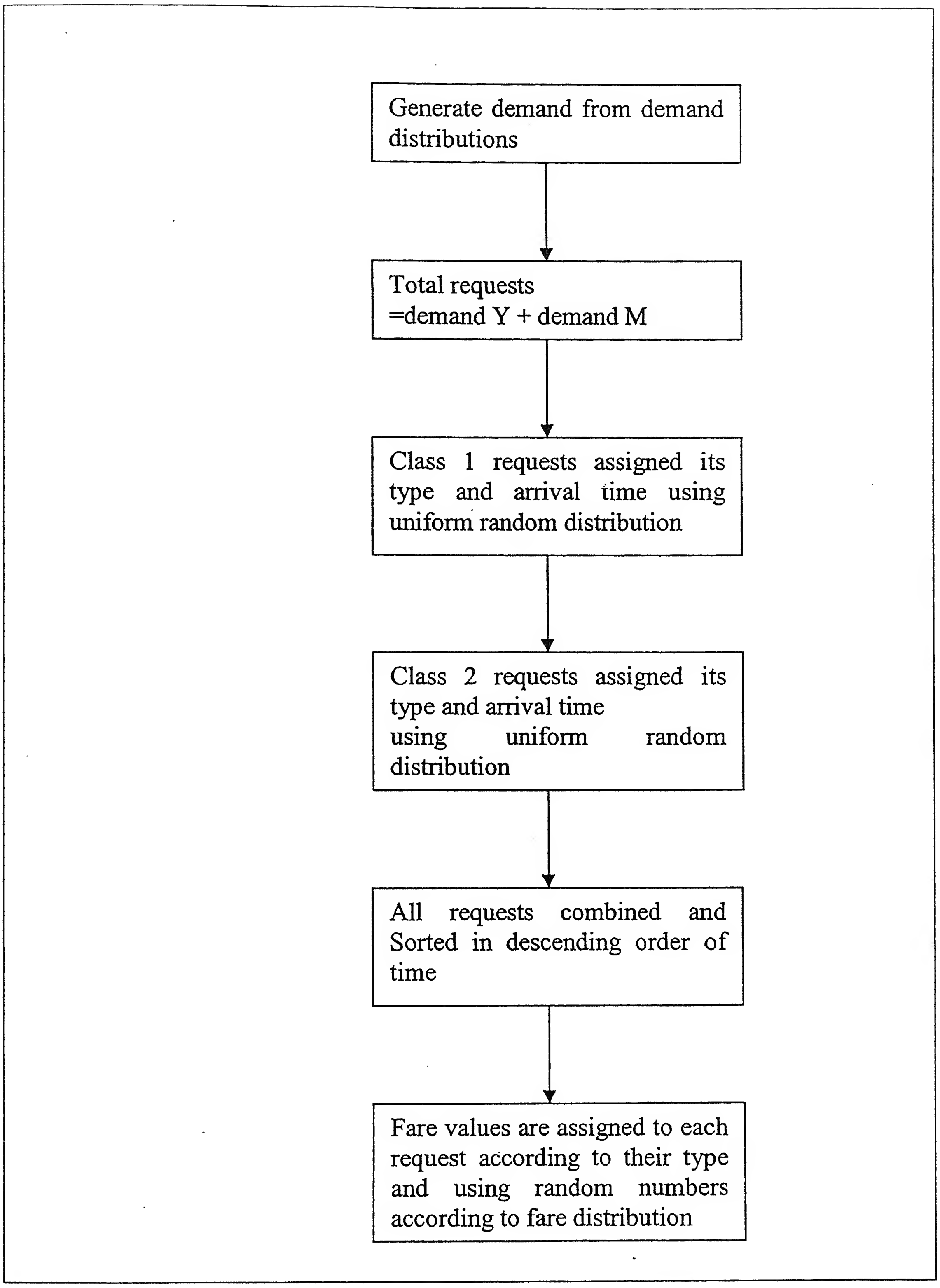


Figure 3.4: Demand generation and Requests attributes

At the end of this process we have complete set of requests with their type, arrival time and actual fare values. This simulates the time and request pattern of the customer for a seat.

The requests are now with all their attributes of type, arrival time and fare value. Now the request of the customer is processed to check if the seat can be allocated to him as per the booking limit for the class requested. As the simulator considers the bookings as nested fare class reservations system, each booking limit is the maximum number of seats that may be sold to a fare class. The booking limit of the highest fare class (class 1 in this case) is the total capacity of the shared cabin. Whereas the booking limit of the lower class 2 is booking limit of class 1 minus the protection level for class 1.

3.1.2.3 Revenue Computation

Class 1 requests are accepted if the seats are available in whole shared cabin. For class 2 the requests are accepted as long as its booking limit is not reached. Once its booking limit is exhausted, any request for class 2 is rejected even if the seats are available in the cabin, which are protected for the higher class 1.The algorithm is depicted in figure. 3.5

```
Let y is no. of seats remaining for Y class
and m is no. of seats remaining for M class

Initially, y = capacity (i.e. booking limit for Y class)and
        m = y - protection level (i.e. booking limit for M class)

If (request type =Y and y not equal to 0 )
        If (y equal to m) then m = m - 1      /*this condition occurs when
                                              y crossed its protection level*/
        y = y-1                               /*every time when y request is
                                              accepted y is minimized by 1*/

elseif (request type = M and m not equal to 0)
        m = m-1                               /*if booking limit of m is not
                                              exhausted minimize m by 1*/
        y = y-1                               /*as y is the total capacity every
                                              time it has to be minimized by 1*/

if (any request accepted) revenue = revenue + fare value
```

Figure 3.5: Algorithm for Revenue Calculation

3.1.2.4 Simulation

The whole process is now replicated 1000 times to get individual demands. New customer requests are generated each time and revenue is then calculated. After 1000 runs the average revenue is calculated. This is shown in the figure 3.6.

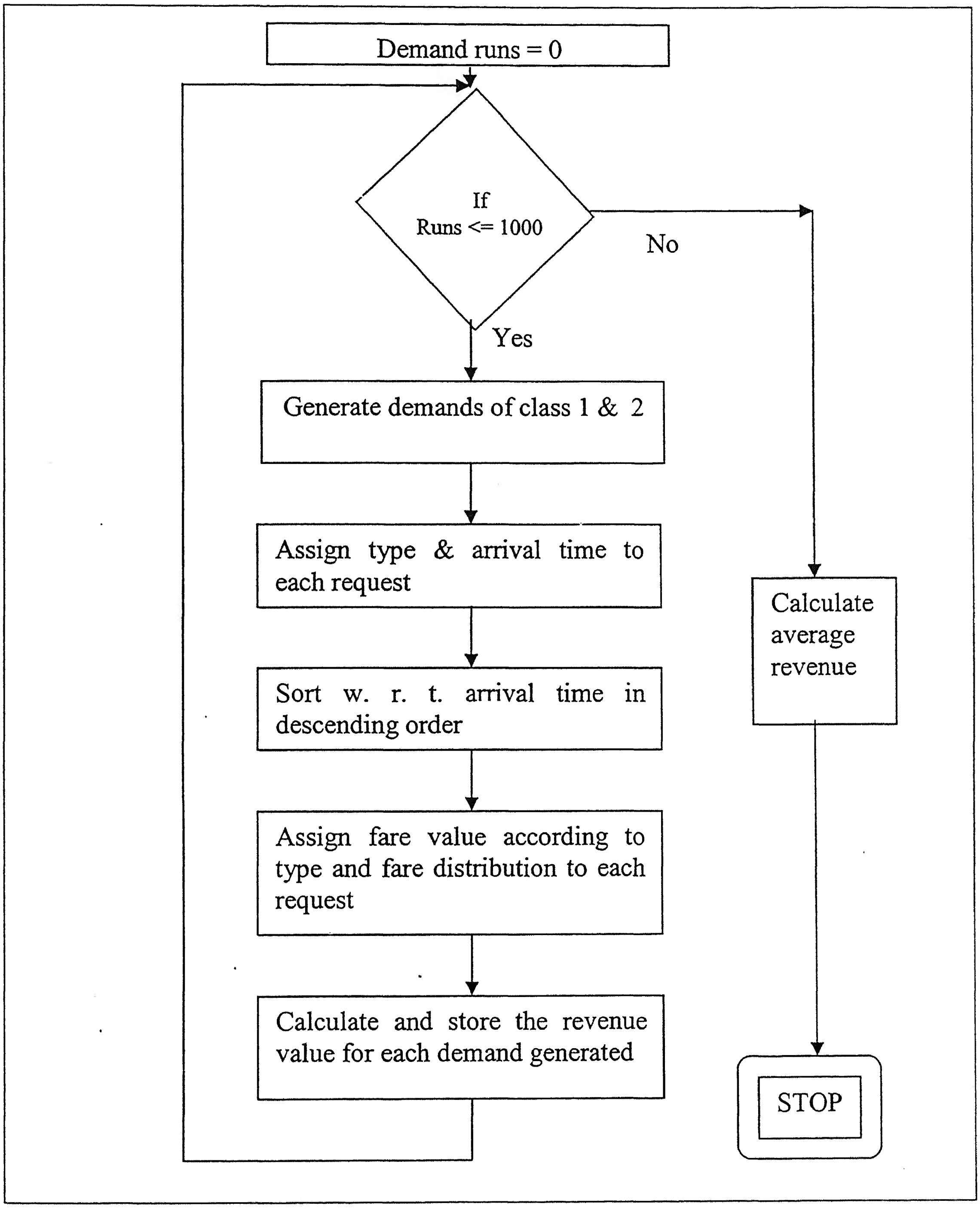


Figure 3.6: General procedure to calculate the average revenue

3.1.2.4.1: Computation of Optimal Protection Level

Above mentioned processes are the building blocks of the simulation program. These blocks can now be used for various purposes according to the need. One of the application of above program is to get the optimal protection level at which the average revenue is maximum. This is computed as follows:

The EMSR protection level is known. One can use this protection level as reference point. Instead of comparing the average revenue for all the values of protection level (from 0 to capacity), one can compute the average revenues for those protection levels which are around the EMSR protection level.

One can calculate the minimum and maximum protection levels with respect to EMSR protection level.

Let P_{min} is minimum protection level

P_{max} is maximum protection level

C is capacity of the coach.

Y_{pro} is protection level for Y class

let the range of protection levels to compare be ± 30 (a number which can be selected).

Taking into consideration the minimum value to be 0 and maximum value to be capacity, the $P_{\min}$ and $P_{\max}$ can be given as,

$$P_{\min} = Y_{pro} - 30 \quad \text{if}\,(P_{\min} < 0)\; P_{\min} = 0$$

$$P_{\max} = Y_{pro} + 30 \quad \text{if}\,(P_{\max} > C)\; P_{\max} = C$$

So the booking limits for M class will be

$$BL_{\min} = C - P_{\min}$$

$$BL_{\max} = C - P_{\max}$$

For each simulation run (1000 replications) consideration of revenue management will compute optimal protection level that is computed as the level

which provide maximum average revenue. But the optimal average revenues are based on the individual random demands and also the random fare values generated for each demand instance in 1000 replications. However if simulation is repeated, it is likely that a different protection level will emerge due to randomness in the data.

To overcome this problem, whole simulation has to be repeated number of times to get a set of the protection level and use this sample data to estimate the protection level. We repeated the simulation 10 times and estimated the optimal protection level as average of sample of 10 protection level is obtained. It will be clear from the figure 3.7.

Let , R_{proy} revenue from protection level given by EMSR

$R_{p\min}$ revenue from protection level $P_{\min}$

$R_{p\max}$ revenue from protection level $P_{\max}$

$Average_i[1]$, $Average_i[2]$, ... $Average_i[r]$,......., $Average_i[k]$

be the average revenues from each of the protection levels.

There will be such 10 averages for each protection level from $P_{\min}$ to $P_{\max}$. Each $Average_i[]$ is the average revenue from 1000 simulation runs.

n	d_1	d_2	$R_{p\min}$					R_{proy}					$R_{p\max}$	
1														
2														
3														
4														
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	.	.	
999	.	.	.	.	.	.	.	.	.	.	.	.	.	
1000														
1			Average1[1]	Average1[2]	.		.	Average1[r]	.	.	.	Average1[k-	Average1[k]	$K_{\max}$ $R_{\max}$
			.	.				.				.	.	.
			.	.				.				.	.	.
10			Average10[1]	Average10[2]	.		.	Average10[r]	.	.	.	Average10[k-	Average10[k]	$K_{\max}$ $R_{\max}$

Figure 3.7: The 10 averages of 1000 simulation runs to get average optimal protection level

From the figure 3.7 it is clear that $R_{\max}$ and hence $K_{\max}$ are the maximum average revenue and optimal protection level out of all other averages. There will be 10 such $R_{\max}$ and $K_{\max}$. And they may or may not be the same. Thus we can get the average of these 10 $K_{\max}$ to get average optimal protection level.

This will be the average optimal protection level for one unit problem from the input file. One can get as many average optimal protection levels as the number of rows in the input file. These protection levels are then used to get the regression model with K (average optimal protection levels) as the dependent variable and the fare distribution, demand distribution, and capacity utilization are controlled or independent variables.

3.1.3 Output File "outfile.xls"

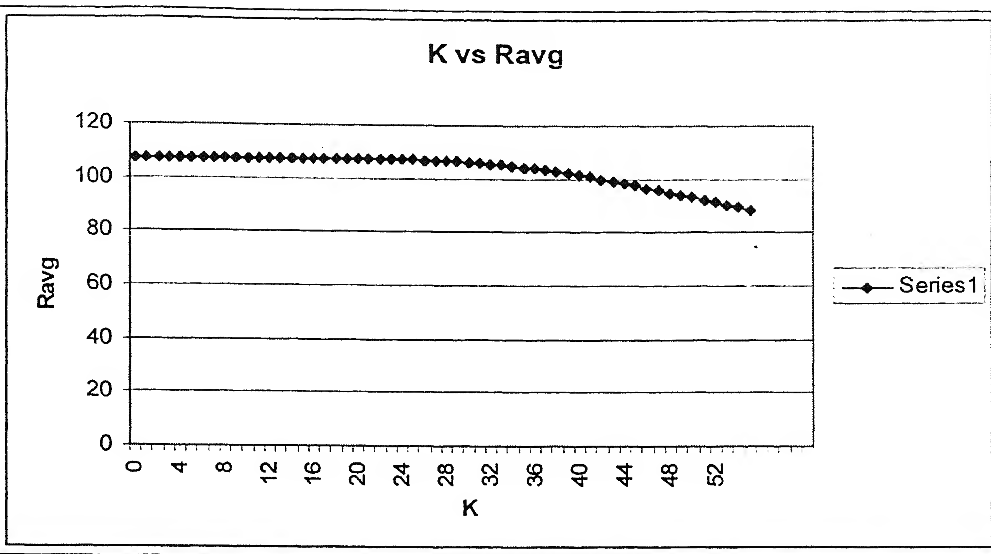
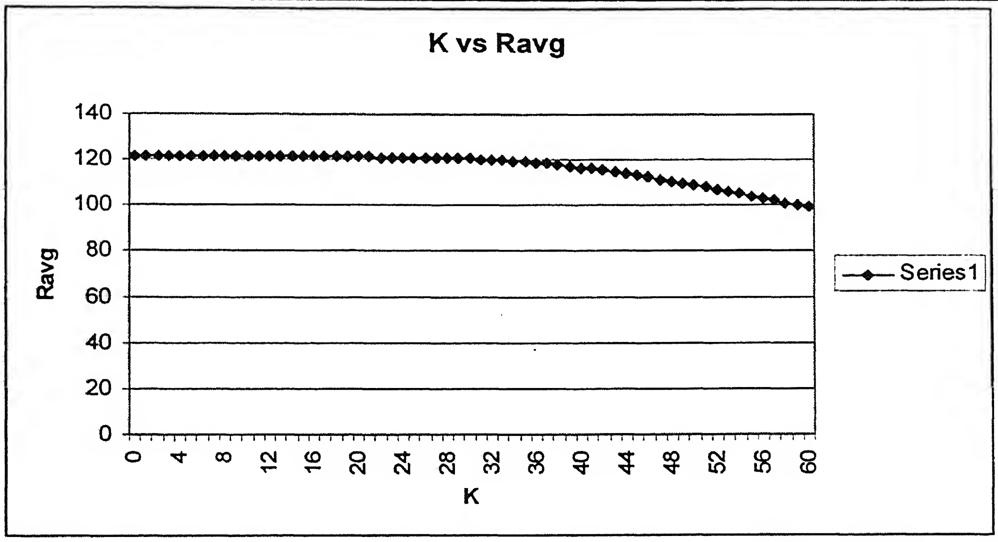
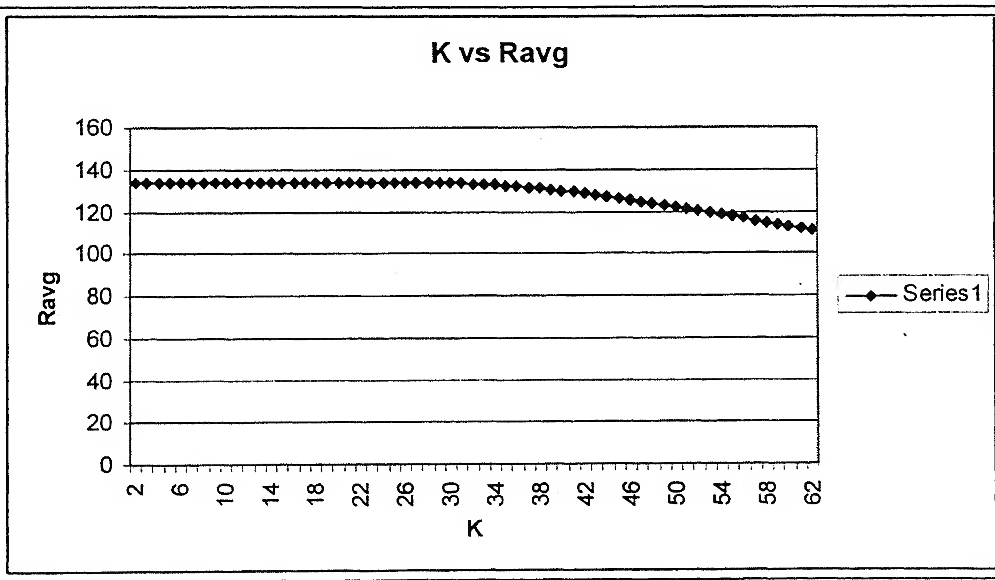
The simulation program generates the excel output file. This file contains the information from input file i.e. the values of parameters and against them the average optimal protection level. It also writes the average revenue from EMSR model and also the optimal average revenues in the output file.

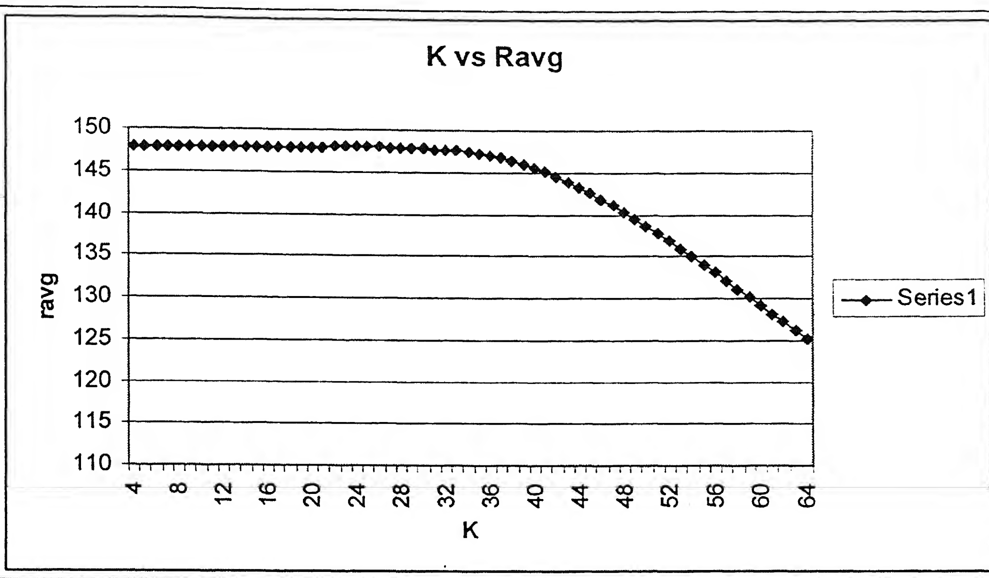
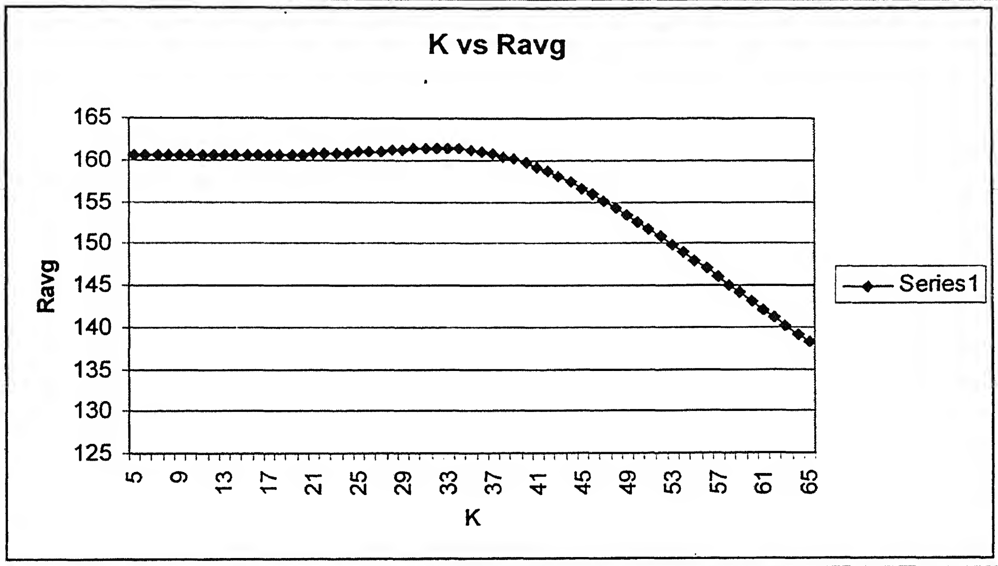
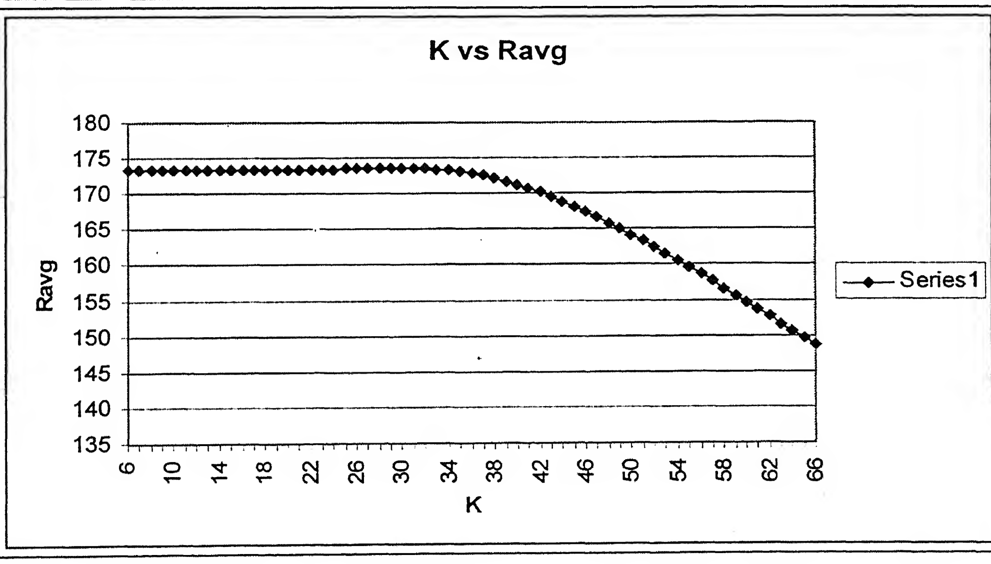
3.2 Simulation Study

From table 3.1 it can be observed that the protection level given by the EMSR model (K_{emsr}) is always higher than the average optimal protection level (K_{star}). The corresponding revenue values (R_{emsr} and R_{star}) are also shown. This data was generated by the simulation code discussed in the previous section where demand distribution is $D_1(30,10)$ and $D_2(75,15)$. μ_{f_1} / μ_{f_2} is changed from 1.5 to 5 with 30% coefficient of variation. Graphs for each μ_{f_1} / μ_{f_2} are plotted to show how average revenue changes with changing protection level.

Table 3.1 values of K_{emsr}, K_{star}, R_{emsr}, R_{star}, for different μ_{f_1} / μ_{f_2} with $D_1(30,10)$ and $D_2(75,15)$

pn	d1	d2	sd1	sd2	f1	f2	s1	s2	K_{emsr}	K_{star}	R_{emsr}	R_{star}
1	30	75	10	15	1.5	1	0.45	0.3	25	4.2	107.4125	107.999
2	30	75	10	15	2	1	0.6	0.3	30	6.5	120.3075	121.3703
3	30	75	10	15	2.5	1	0.75	0.3	32	11.5	134.1296	134.9987
4	30	75	10	15	3	1	0.9	0.3	34	18.1	147.7954	148.5557
5	30	75	10	15	3.5	1	1.05	0.3	35	26.8	161.5751	162.1748
6	30	75	10	15	4	1	1.2	0.3	36	32	175.9367	176.2852
7	30	75	10	15	5	1	1.5	0.3	38	34.8	203.7466	204.0369

Problem 1 $\mu_{f_1} / \mu_{f_2} = 1.5$	
Problem 2 $\mu_{f_1} / \mu_{f_2} = 2$	
Problem 3 $\mu_{f_1} / \mu_{f_2} = 2.5$	

Problem 4 $\mu_{f_1}/\mu_{f_2}=3$	
Problem 5 $\mu_{f_1}/\mu_{f_2}=3.5$	
Problem 6 $\mu_{f_1}/\mu_{f_2}=4$	

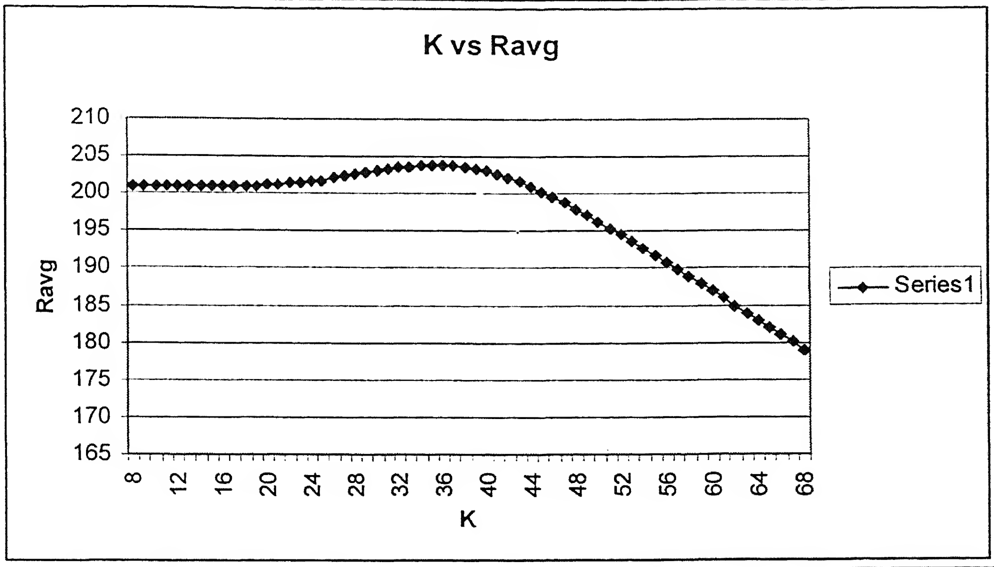
Problem 7 $\mu_{f_1}/\mu_{f_2} = 5$	

Figure 3.8: Relationship between protection level K and average revenue R_{avg} for different fare ratios

The purpose of these graphs is to show change in the average revenue with the protection level. As can be seen from the graphs, the average revenue is low and constant for low protection levels. Then it increases slowly as protection level is increased. It is maximum at a certain protection level and then starts decreasing.

This behavior of average revenue with respect to protection level emphasizes the importance of choosing the appropriate protection level to maximize the revenue. The reason behind this shape of the graph is briefly discussed below.

Let us assume a high demand scenario where demand is greater than supply. There is a high demand for airline seats from both the classes, class 1 and 2. With low protection levels, more of class 2 customers (low paying) will occupy the shared cabin and once the capacity of the flight exhausts the requests from class 1 customers (high paying), coming late, will be denied. This leads to less revenue to the airline. Now if the protection level is increased, more of the requests from class 2 customers will be rejected once their booking limits are exhausted. This will maximize the revenue only when all the protected seats for class 1 are filled. When more seats are protected for class 1 customers, and it happens that demand is less for these seats, these seats will go empty. It will result in a loss to airline. The figure 3.8 shows a linear relationship for this phenomena. Higher the protection level (after a certain point), less will be the average revenue.

This is true when there is very little or no overlap between the fare distribution of the two classes i.e. μ_{f_1}/μ_{f_2} high. But when the mean fares of the two fare classes are very close i.e. μ_{f_1}/μ_{f_2} low, behavior of the graph changes. There is not much difference in the actual fares paid by the customers of the two classes. So at low protection levels, the average revenue remains nearly constant. The problem arises only when the protection level is increased because in that case the requests from class 2 customers will be rejected for a future class 1 request which may be of nearly same fare value. It may also happen that the future request from class 1 customer may be of low fare value than the currently denied class 1 customer request. Thus the seat either remains empty if there is no show of class 1 customer or it generates a little higher revenue. Practically speaking high protection levels is not the best strategy when μ_{f_1}/μ_{f_2} is very low.

Thus the problem of deciding on the protection level becomes crucial when the μ_{f_1}/μ_{f_2} is not so high and not so low. This is the case on which majority of fare structures currently present in the market are based. Airlines have to make a tradeoff on protection levels to accommodate for the high paying class2 customers instead of letting the seats to go empty. The graph of average revenue is increasing w.r.t. protection level upto certain point and then starts moving downwards. It can be further observed that the slope in its downward direction is increasing with increase in the μ_{f_1}/μ_{f_2} ratio.

From the table 3.1, it can be further observed that the average optimal protection level is always less than the EMSR protection level. This observation implies that one can improve on revenue if the protection levels are kept lower than that given by EMSR. This leads to need of developing model which finds optimal protection levels.

Once the protection levels are set and booking limit for class 2 is exhausted, no more requests for class 2 are accepted even if seats are available in the cabin, which are protected for class 1 customers. Also if the class 1 customer request arrives, and seats are available in the cabin it will be readily accepted irrespective of what he is paying. Basically, the lower optimal protection levels are meant to accommodate

some of the class 2 (low mean fare) customers who pay more than class 1 (high mean fare) customers.

From table 3.1, it can further be observed that as the μ_{f_1} / μ_{f_2} ratio increases the optimal protection level is shifting towards EMSR protection level. This is because the number of class 2 customers paying higher than class 1 customers is decreasing. Thus there is no need to keep extra seats for these class 2 customers. In other words the protection level can be increased if the overlap of fare distribution decreases.

Observations in this section leads to a more dedicated exploration of relationship between optimal protection level and fare structure. This is studied in the next section. In the same section we will study how much revenue can be gained by using the strategy of low protection levels.

3.3 Study of Optimal Protection Level For Different Demand Distributions

A comparative study was done between the performance of EMSR protection levels (K_{emsr}) and optimal protection levels (K_{star}) over different values of for μ_{f_1} / μ_{f_2} different demand distributions. The purpose is to study the behavior of K_{emsr} and K_{star} over the range of μ_{f_1} / μ_{f_2} for different demand distributions and to study how much improvement can be made over traditional EMSR model measured by

$$\%improvemen\,t = \frac{(R_{star} - R_{emsr})}{R_{emsr}} * 100$$

Following three typical examples were considered to do the study where the demand distribution is considered to be of high demand i.e. capacity utilization ranging from 0.9 to 1.15. Capacity is taken to be 100.

Example 1

In this example, the demand distribution is $D_1(20,7)$ and $D_2(70,28)$ with capacity utilization of 0.9.

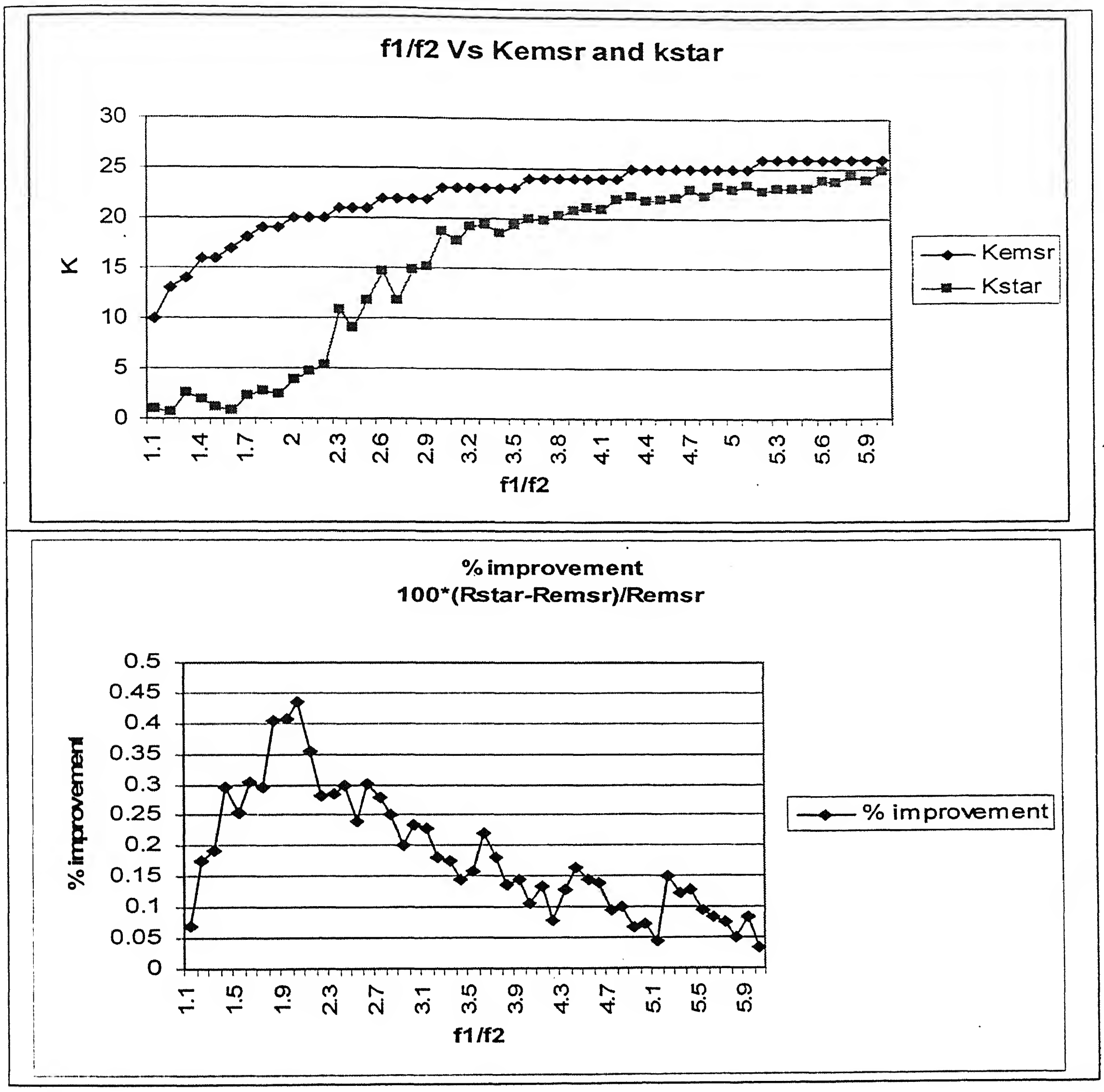


Figure 3.9: Graph of K and %improvement against μ_{f_1}/μ_{f_2} for $D_1(20,7)$ and $D_2(70,28)$

For a demand ratio μ_{d_1}/μ_{d_2} of 20/70, the optimal protection level is less than 5 over the range of μ_{f_1}/μ_{f_2} ratio from 1.1 to 2.3 and then it jumps over to 20 for μ_{f_1}/μ_{f_2} at 4. It then increases steadily to catch up with the EMSR protection level for μ_{f_1}/μ_{f_2} to be greater than 4.5.

The percent improvement reaches to its maximum of 0.45 at μ_{f_1}/μ_{f_2} to be around 2.2. In all, the percent improvement above 0.3 % can be seen in the μ_{f_1}/μ_{f_2} range of 2 to 3.

Example 2

In this example, the demand distribution is $D_1(30,11)$ and $D_2(80,32)$ with capacity utilization of 1.1.

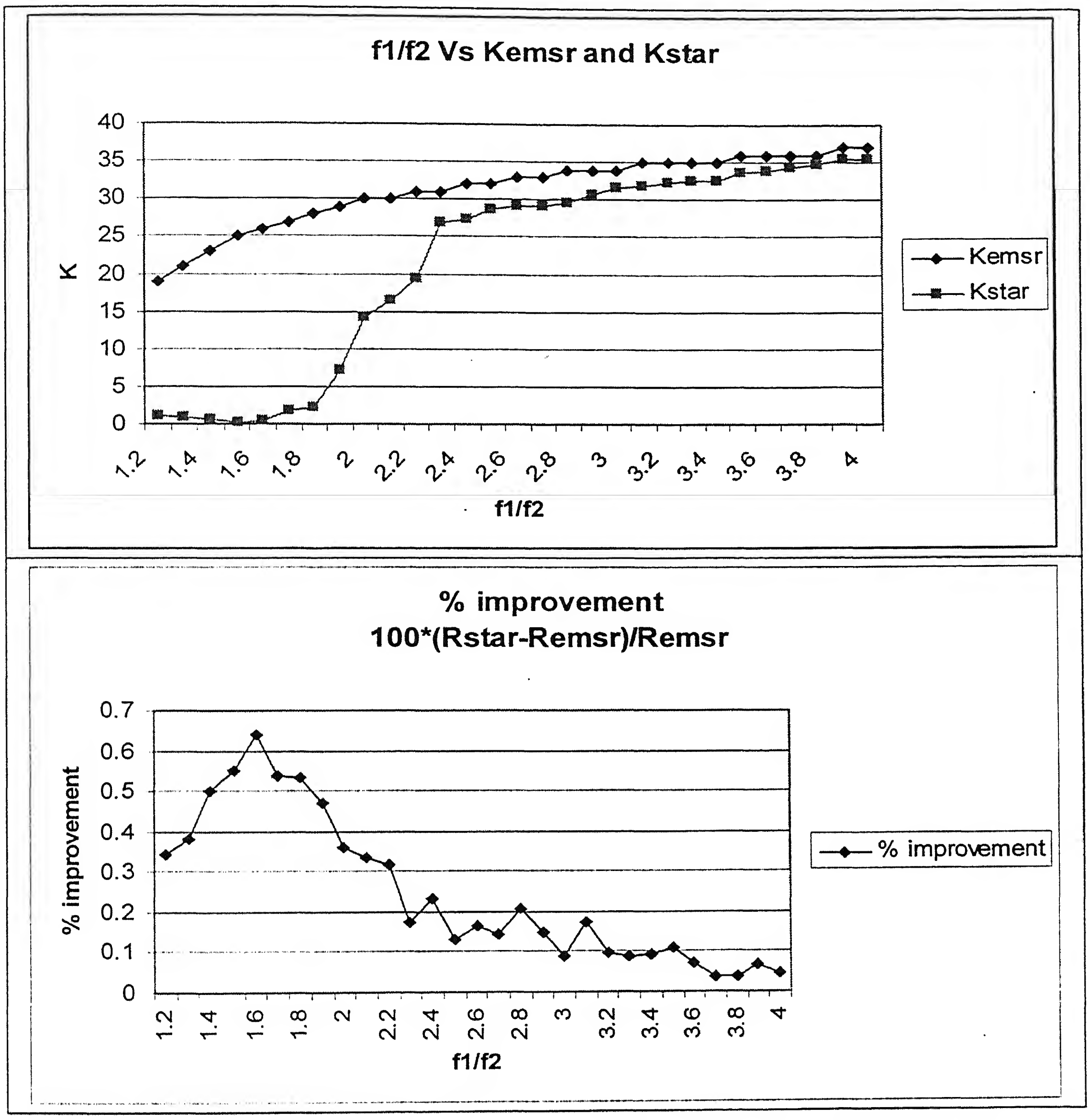


Figure 3.10: Graph of K and %improvement against μ_{f_1}/μ_{f_2} for $D_1(30,11)$ and $D_2(80,32)$

For a demand ratio (μ_{d_1}/μ_{d_2}) of 30/80, the optimal protection level is less than 5 over the range of μ_{f_1}/μ_{f_2} ratio from 1.1 to 2 and then it jumps over to 25 for μ_{f_1}/μ_{f_2} at 2.3. It then increases steadily to catch up with the EMSR protection level for f1/f2 to be greater than 2.5

The percent improvement reaches to its maximum of 0.65 at μ_{f_1}/μ_{f_2} to be around 1.5. In all, the percent improvement above 0.3 % can be seen in the μ_{f_1}/μ_{f_2} range of 1.2 to 2.2.

Example 3

In this example, the demand distribution is $D_1(40,14)$ and $D_2(75,30)$ with capacity utilization of 1.15

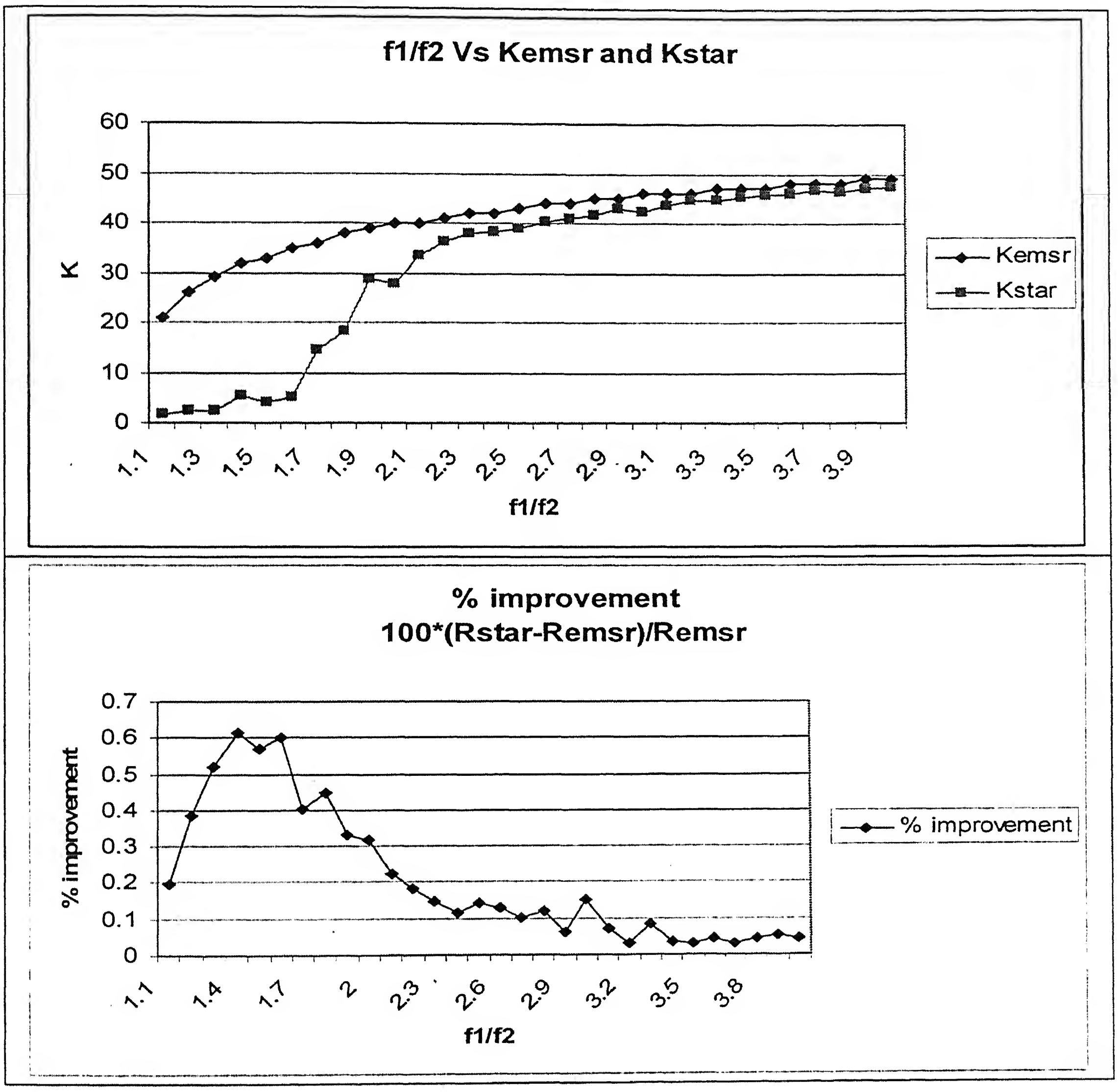


Figure 3.11: Graph of K and %improvement against μ_{f_1}/μ_{f_2} for $D_1(40,14)$ and $D_2(75,30)$

For a demand ratio (μ_{d_1}/μ_{d_2}) of 40/75, the optimal protection level is less than 10 over the range of μ_{f_1}/μ_{f_2} ratio from 1.1 to 1.7 and then it jumps over to 40

for μ_{f_1} / μ_{f_2} at 2.3. It then increases steadily to catch up with the EMSR protection level for μ_{f_1} / μ_{f_2} to be greater than 2.5

The percent improvement reaches to its maximum of 0.6 at μ_{f_1} / μ_{f_2} to be around 1.5. In all, the percent improvement above 0.3 % can be seen in the μ_{f_1} / μ_{f_2} range of 1.2 to 2.

Studying these examples, it is clear that there is less than 0.5% improvement when capacity utilization is less than 1 and more than 0.5% when it is more than 1. This improvement is also for a specific range of μ_{f_1} / μ_{f_2} which too depends upon the demand distribution. Typically % improvement more than 0.3% can be made in the range of μ_{f_1} / μ_{f_2} from 1.2 to 2.2.

This study also reveals that the capacity utilization plays important role in deciding the amount of revenue gain. In the next section we will study the effects of various parameters on the optimal protection level by means of Design of Experiments.

3.4 Experimental Design

Experiments are carried out by investigators in all fields of study either to discover something about a particular process or to compare the effect of several conditions on some phenomena. In the current study, experiments were carried out to compare the effect of several conditions like factor ratio μ_{f_1} / μ_{f_2}, demand ratio μ_{d_1} / μ_{d_2}, capacity utilization, coefficient of variation etc. on the protection level.

The simulation study conducted in previous section revealed that the EMSR protection levels are not optimal at low μ_{f_1} / μ_{f_2} ratio. The factors affecting the protection level are thus the parameters of fare distribution and demand distribution. In this section the effect of several of these factors on the protection level are studied using Design of Experiment.

It is not necessary that all the factors considered are important. To find out what are the significant factors and what factors can be ignored we design the

experiments and use statistical inference techniques. Analysis of Variance ANOVA is probably the most useful technique in the field of statistical inference. And hence one can have a certain level of confidence in his decisions of selecting the significant factors and ignoring insignificant ones.

After getting acquainted with the significant factors affecting the protection level, the next step is to explore the nature of relationship between them. It will obviously be helpful if one can predict the protection levels in advance considering the relationship. This relationship is in a functional form can be established using Regression.

3.4.1 Design of Experiments

In any design of experiments, the set of factors and their levels must be chosen carefully so that one can get the maximum information of the experimented phenomena. In this study it is decided to first consider the factors associated with fare distribution and conduct the experiment. The fare structures are within the control of the airlines and hence can be treated as controlled variable. The demand distributions may be influenced by the fare demands, but are not controlled by the airlines and hence are blocking variables. Initially we designed experiments with fare structure variable and treating demand variable as part of the random noise.

The factors chosen for the experiment are μ_{f_1}/μ_{f_2}, σ_{f_1}/μ_{f_1} and σ_{f_2}/μ_{f_2}. It was decided to conduct a three level full factorial experimental design. Factorial designs are used to study the joint effect of these factors on the response variable i.e. protection level. The low value, medium value and high value of the three factors was decided as the levels for each of the factor. The values are shown in the table 3.2.

Table 3.2: DOE for μ_{f_1}/μ_{f_2}, σ_{f_1}/μ_{f_1} and σ_{f_2}/μ_{f_2}

μ_{f_1}/μ_{f_2}	σ_{f_1}/μ_{f_1}	σ_{f_2}/μ_{f_2}
1.25	0.1	0.1
2.5	0.2	0.2
3.75	0.3	0.3

A specific demand distribution e.g. $D_1(40,14)$ and $D_2(75,30)$ was considered to conduct the experiment. Due to three factors and three levels, 27 experiments were

conducted. Each experiment was a unit problem to the simulation program. Each experiment was replicated 10 times. The output of the program stored in the output file was used for statistical analysis.

3.4.2 Analysis of Variance

After performing the experiment the next step is to analyze the results statistically. Analysis of Variance (ANOVA) is a very useful statistical inference technique and is used in the current study to decide the significant and insignificant factors. Mathematica software was being used to conduct ANOVA which considers the standard 95% confidence level for the test. So if the P value is below 0.05 then our factor is significant otherwise it is insignificant.

In the design following factors were considered:

$\alpha = \mu_{f_1} / \mu_{f_2}$

$\beta = \sigma_{f_1} / \mu_{f_1}$ and

$\gamma = \sigma_{f_2} / \mu_{f_2}$

ANOVA conducted on above full factorial design generated by Mathematica is as shown below in the table form. The rightmost column shows the P-value.

The P-value for alpha (μ_{f_1} / μ_{f_2}) is coming out to be zero which signifies that μ_{f_1} / μ_{f_2} is the most significant factor in the study. All other factors and their interactions can be considered insignificant as their P-value is greater than 0.05.

Table 3.3: ANOVA result generated by Mathematica for μ_{f_1}/μ_{f_2}, σ_{f_1}/μ_{f_1} and σ_{f_2}/μ_{f_2}

ANOVA→

	DF	SumOfSq	MeanSq	FRatio	PValue
α	2	102492.	51246.1	16706.2	0.
β	2	8.86667	4.43333	1.44526	0.237702
γ	2	4.35556	2.17778	0.709954	0.492684
$\alpha\beta$	4	26.2444	6.56111	2.13892	0.0766359
$\alpha\gamma$	4	7.08889	1.77222	0.577743	0.679041
$\beta\gamma$	4	7.57778	1.89444	0.617588	0.650381
$\alpha\beta\gamma$	8	40.9778	5.12222	1.66984	0.106338
Error	243	745.4	3.06749		
Total	269	103333.			

On the same lines the design of experiment was performed to check whether the ratio of standard deviations of two fare distributions have any effect on optimal protection level. The factors chosen for the experiment were μ_{f_1}/μ_{f_2} and $\sigma_{f_1}/\sigma_{f_2}$. It was decided to conduct a three level full factorial experimental design. The low value, medium value and high value of the two factors is as shown in the table 3.4.

Table 3.4 DOE for μ_{f_1}/μ_{f_2} and $\sigma_{f_1}/\sigma_{f_2}$

μ_{f_1}/μ_{f_2}	$\sigma_{f_1}/\sigma_{f_2}$
1.25	1
2.5	2
3.75	3

The demand distribution is kept same as earlier e. g. $D_1(40,14)$ and $D_2(75,30)$. Here first σ_{f_1} values were fixed at 0.125, 0.25 and 0.375 and σ_{f_2} values changed accordingly to follow the scheme of $\sigma_{f_1}/\sigma_{f_2}$ = 1, 2, 3.So a total of 9 experiments were performed. Each experiment was replicated 10 times.

In the design following factors were considered:

$\alpha = \mu_{f_1}/\mu_{f_2}$

$\beta = \sigma_{f_1}/\sigma_{f_2}$

ANOVA conducted on above full factorial design generated by Mathematica is as shown below in the table 3.5. The rightmost column shows the P-value.

The P-value for alpha (μ_{f_1}/μ_{f_2}) is coming out to be zero which signifies that μ_{f_1}/μ_{f_2} is the most significant factor in the study. All other factors and their interactions can be considered insignificant as their P-value is greater than 0.05.

Table 3.5: ANOVA result generated by Mathematica for μ_{f_1}/μ_{f_2} and $\sigma_{f_1}/\sigma_{f_2}$

ANOVA→

	DF	SumOfSq	MeanSq	FRatio	PValue
α	2	32407.8	16203.9	4227.1	0.
β	2	1.35556	0.677778	0.176812	0.83826
$\alpha\beta$	4	13.5111	3.37778	0.881159	0.479052
Error	81	310.5	3.83333		
Total	89	32733.1			

From the above two experiments it was concluded that μ_{f_1}/μ_{f_2} is the most significant factor in deciding the optimal protection level. In the next step demand distributions were considered as blocking variables and their effects on optimal protection level were studied.

3.4.3 Design of Experiments Considering Demand Distributions

After studying the effect of μ_{f_1}/μ_{f_2} on K_{star} for a specific demand distribution, we will explore the effect of demand distribution on K_{star}. The demand distribution has four parameters namely μ_{d_1}, σ_{d_1} and μ_{d_2}, σ_{d_2}. The current study deals with the segregated market i.e. class 1 and class2 customers. So μ_{d_1}/μ_{d_2} can be considered as one of the important factors. σ_{d_1}/μ_{d_1} and σ_{d_2}/μ_{d_2} are also considered as other two factors.

As the fare structure (controlled variable) are considered, we are concerned with the overlapping area of the two fare distributions (f_1 and f_2) as shown in figure 3.12. This area signifies the probability of class 1 customers paying low fares than the

class 2 customers. The traditional EMSR model assumes only the mean fares paid by the customers and not their spread. We denote this area by Pm.

Fare structure is represented by system controlled value Pm as a factor considered in the study. The last factor considered is capacity utilization ($\mu_{d_1} + \mu_{d_2}$)/C, which may have impact on the protection level.

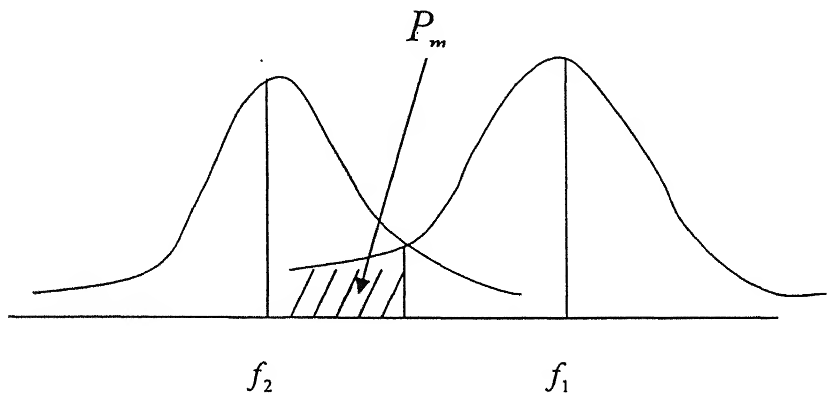


Figure 3.12: Probability that class 1 customers paying less than class 2 customers

In this study we consider the five factors as follows:

1. $(\mu_{d_1} + \mu_{d_2})/C$: C_u

2. μ_{d_1} / μ_{d_2} : R_d

3. P_m : P_m

4. σ_{d_1} / μ_{d_1} : V_1

5. σ_{d_2} / μ_{d_2} : V_2

Each factor is considered as three levels, the high level, middle level and low level. The values are shown in the table 3.6

Table 3.6: DOE for C_u, R_d, P_m, V_1, V_2

C_u	R_d	P_m	V_1	V_2
0.8	0.1	0.057	0.200	0.200
1	0.29	0.118	0.250	0.250
1.2	0.6	0.335	0.333	0.333

There are five factors and each factor was studied at three levels so a total of 243 experiments were performed. Each experiment was replicated 10 times.

From the ANOVA table 3.7 most of the factors and their interactions are found to be significant whereas some interactions are insignificant. Following is the list of significant factors and significant interactions.

C_u	R_d	P_m	V_1	V_2	
$C_u\,R_d$	$C_u\,P_m$	$C_u\,V_1$	$C_u\,V_2$	$R_d\,P_m$	$R_d\,V_1$
$R_d\,V_2$	$P_m\,V_1$	$P_m\,V_2$	$V_1\,V_2$	$C_u\,R_d\,P_m$	$C_u\,R_d\,V_1$
$C_u\,R_d\,V_2$	$C_u\,P_m\,V_1$	$C_u\,P_m\,V_2$	$R_d\,P_m\,V_1$	$R_d\,P_m\,V_2$	$R_d\,V_1\,V_2$
$P_m\,V_1\,V_2$	$C_u\,R_d\,P_m\,V_1$	$R_d\,P_m\,V_1\,V_2$	$C_u\,R_d\,P_m\,V_1\,V_2$		

3.5 Regression Analysis and Model Development

After having sufficient knowledge of the significant factors and significant interactions from ANOVA, the next step is to conduct Regression Analysis. Regression analysis reveals the nature of relationship between the dependent variable and various independent variables. In the current study, optimal protection level is the dependent variable whereas other variables derived from demand distributions and fare distributions are independent variables.

A simple hypothetical model is developed first based on the prior knowledge of various factors affecting the dependent variable and unknown coefficients. Regression analysis finds the value of these unknown coefficients based on the available data and fits the curve so that the error of fit is minimum. The mathematical

Table 3.7 ANOVA result generated by Mathematica for C_u, R_d, P_m, V_1, V_2

		DF	SumOfSq	MeanSq	FRatio	PValue
	A	2	5057.27	2528.64	2934.59	0.
	B	2	29370.1	14685.	17042.6	0.
	C	2	10282.5	5141.25	5966.64	0.
	D	2	4157.13	2078.57	2412.26	0.
	e	2	32.126	16.063	18.6418	2.93804×10^{-8}
	A B	4	893.193	223.298	259.147	0.
	A C	4	4445.74	1111.43	1289.87	0.
	A D	4	622.383	155.596	180.575	0.
	A e	4	536.879	134.22	155.768	0.
	B C	4	2310.65	577.663	670.402	0.
	B D	4	571.688	142.922	165.867	0.
	B e	4	52.0342	13.0085	15.097	4.89232×10^{-11}
	C D	4	469.424	117.356	136.196	0.
	C e	4	151.02	37.7549	43.8162	0.
	D e	4	11.6665	2.91663	3.38487	0.0102068
	A B C	8	1382.75	172.844	200.593	0.
ANOVA →	A B D	8	241.055	30.1318	34.9693	0.
	A B e	8	259.079	32.3848	37.5839	0.
	A C D	8	1064.35	133.044	154.403	0.
	A C e	8	243.716	30.4646	35.3554	0.
	A D e	8	0.0025	0.0003125	0.000362669	1.
	B C D	8	95.3158	11.9145	13.8273	1.11022×10^{-16}
	B C e	8	60.0476	7.50595	8.71096	1.87812×10^{-10}
	B D e	8	38.5293	4.81616	5.58936	1.6499×10^{-6}
	C D e	8	105.152	13.144	15.2542	0.
	A B C D	16	245.53	15.3456	17.8093	0.
	A B C e	16	129.054	8.06588	9.36079	0.
	A B D e	16	-3.15252×10^{16}	-1.97032×10^{15}	-2.28664×10^{15}	1
	A C D e	16	310.466	19.4041	22.5193	0.
	B C D e	16	85.3579	5.33487	6.19133	1.74492×10^{-11}
	A B C D e	32	222.806	6.9627	8.08051	0.
	Error	243	209.385	0.861667		
	Total	485	63776.9			

model which is developed is the regressed model which can predict the value of dependent variable.

In the current study the optimal protection level is the dependent variable. This variable is to be predicted beforehand by means of its relationship with known or estimated factors. A mathematical model is to be developed by means of regression analysis. The data required to perform regression analysis is generated from the simulation program. For various combinations of independent variables, optimal protection levels were generated from the simulation program.

Using this data regression analysis provides the estimates of the coefficients of various terms used in the model. These estimates when used in the hypothetical model gives us a mathematical model which will predict the optimal protection level under given set of conditions.

In the current study SPSS statistical software was used to perform the regression analysis. The data generated by the simulation program is then fed in SPSS statistical software along with the multivariate linear regression model. The starting value of various coefficients is set to "0". SPSS gives out the estimates of various coefficients along with the correlation coefficient and other statistics.

Following two models were developed during the study. The first model gets the optimal protection level considering five cases of capacity utilizations separately. The second model considers the capacity utilization as one of the independent variable. Both the models are as discussed below.

3.5.1 Model 1: Considering Five Cases of Capacity Utilization

The first model was developed as a combination of five cases. The five cases include the capacity utilization values ranging from 0.8, 0.9, 1, 1.1 and 1.2. A separate regression analysis was done for each of the capacity utilization.

As discussed earlier the results of ANOVA stated all the factors and their interactions to be significant. So a multivariate linear model common to all five cases was tested. The model is discussed below:

The four independent variables in the suggested models are:

1. μ_{d_1} / μ_{d_2} : R_d

2. P_m : P_m

3. σ_{d_1} / μ_{d_1} : V_1

4. σ_{d_2} / μ_{d_2} : V_2

The model is:

$$K_{star} = \alpha_0 + \alpha_1 R_d + \alpha_2 P_m + \alpha_3 V_1 + \alpha_4 V_2 + \alpha_5 R_d P_m + \alpha_6 R_d V_1 + \alpha_7 R_d V_2 + \alpha_8 P_m V_1 + \alpha_9 P_m V_2 + \alpha_{10} V_1 V_2 + \alpha_{11} R_d P_m V_1 + \alpha_{12} R_d P_m V_2 + \alpha_{13} R_d V_1 V_2 + \alpha_{14} P_m V_1 V_2 + \alpha_{15} R_d P_m V_1 V_2$$

This multivariate linear model was fed in as hypothetical model for Regression Analysis in the SPSS software along with the parameter values from α_0 to α_{14} with their initial values set at "0". For each capacity utilization, a different data set was used. The mathematical model generated for each of this case is as shown below:

Case 1: for capacity utilization = 0.8

The estimates of the coefficients α_0 to α_{14} can be found from the regression table C-0.8 in Appendix C. R squared = 0.89726

$$K_{star} = 32.369 - 63.942 R_d - 413.927 P_m - 158.601 V_1 - 111.024 V_2 + 1215.418 R_d P_m + 465.873 R_d V_1 + 464.074 V_2 R_d + 1550.644 P_m V_1 + 1769.674 P_m V_2 + 591.128 V_1 V_2 - 4208.913 R_d P_m V_1 - 5633.053 R_d P_m V_2 - 2134.737 R_d V_1 V_2 + - 6752.903 P_m V_1 V_2 + 19459.733 R_d P_m V_1 V_2$$

$$K_{star} \geq 0$$

Case 2: for capacity utilization = 0.9

The estimates of the coefficients α_0 to α_{14} can be found from the regression table C-0.9 in Appendix C. R squared = .84555

$$K_{star} = 79.668 - 239.831 R_d - 92.640 P_m - 308.860 V_1 - 261.909 V_2 + 661.891 R_d P_m + 1013.448 R_d V_1 + 1175.409 V_2 R_d + 62.573 P_m V_1 + 477.268 P_m V_2 + 1065.252 V_1 V_2 - 1171.706 R_d P_m V_1 - 3875.597 R_d P_m V_2 - 4285.613 R_d V_1 V_2 - 736.029 P_m V_1 V_2 + 8199.035 R_d P_m V_1 V_2$$

$K_{star} \geq 0$

Case 3: for capacity utilization = 1.0

The estimates of the coefficients α_0 to α_{14} can be found from the regression table C-1.0 in Appendix C. R squared = .84184

$$K_{star} = -61.025 + 400.439 R_d - 47.670 P_m + 256.969 V_1 + 207.475 V_2$$
$$+ 192.866 R_d P_m - 1454.846 R_d V_1 - 973.820 R_d V_2 + 30.622 P_m V_1$$
$$+ 739.024 P_m V_2 - 807.830 V_1 V_2 - 344.488 R_d P_m V_1 - 2591.690 R_d P_m V_2$$
$$+ 4169.013 R_d V_1 V_2 - 2517.048 P_m V_1 V_2 + 6612.436 R_d P_m V_1 V_2$$
$$K_{star} \geq 0$$

Case 4: for capacity utilization = 1.1

The estimates of the coefficients α_0 to α_{14} can be found from the regression table C-1.1 in Appendix C. R squared = .94390

$$K_{star} = -15.101 + 144.153 R_d + 783.703 P_m + 80.340 V_1 + 95.276 V_2$$
$$- 3663.805 R_d P_m - 363.667 R_d V_1 - 351.046 R_d V_2 - 3029.747 P_m V_1$$
$$- 3742.979 P_m V_2 - 382.289 V_1 V_2 + 13618.585 R_d P_m V_1 + 17053.957 R_d P_m V_2$$
$$+ 1644.847 R_d V_1 V_2 + 14537.962 P_m V_1 V_2 - 67798.832 R_d P_m V_1 V_2$$
$$K_{star} \geq 0$$

Case 5: for capacity utilization = 1.2

The estimates of the coefficients α_0 to α_{14} can be found from the regression table C-1.1 in Appendix C. R squared = .96764

$$K_{star} = -41.403 + 224.346 R_d + 886.667 P_m + 156.696 V_1 + 160.086 V_2$$
$$- 393.700 R_d P_m - 530.328 R_d V_1 - 514.344 R_d V_2 - 3053.097 P_m V_1$$
$$- 3105.923 P_m V_2 - 520.101 V_1 V_2 - 1060.463 R_d P_m V_1 + 2928.702 R_d P_m V_2$$
$$+ 1861.766 R_d V_1 V_2 + 10516.763 P_m V_1 V_2 - 5268.755 R_d P_m V_1 V_2$$
$$K_{star} \geq 0$$

3.5.2 Model 2: Considering Capacity Utilization as One of the Factors

As discussed earlier the results of ANOVA stated most of the factors and their interactions were significant. The insignificant interactions were excluded from the study. So a multivariate linear model was suggested considering only the significant factors and significant interactions.

The five independent variables in the suggested models are:

1. $(\mu_{d_1} + \mu_{d_2})/C$: C_u

2. μ_{d_1} / μ_{d_2} : R_d

3. P_m : P_m

4. σ_{d_1} / μ_{d_1} : V_1

5. σ_{d_2} / μ_{d_2} : V_2

The model is

$$K_{star} = \alpha_0 + \alpha_1 C_u + \alpha_2 R_d + \alpha_3 P_m + \alpha_4 V_1 + \alpha_5 V_2 + \alpha_6 C_u R_d + \alpha_7 C_u P_m + \alpha_8 C_u V_1 + \alpha_9 C_u V_2 + \alpha_{10} R_d P_m + \alpha_{11} R_d V_1 + \alpha_{12} R_d V_2 + \alpha_{13} P_m V_1 + \alpha_{14} P_m V_2 + \alpha_{15} V_1 V_2 + \alpha_{16} C_u R_d P_m + \alpha_{17} C_u R_d V_1 + \alpha_{18} C_u R_d V_2 + \alpha_{19} C_u P_m V_1 + \alpha_{20} C_u P_m V_2 + \alpha_{21} R_d P_m V_1 + \alpha_{22} R_d P_m V_2 + \alpha_{23} R_d V_1 V_2 + \alpha_{24} P_m V_1 V_2 + \alpha_{25} C_u R_d P_m V_1 + \alpha_{26} R_d P_m V_1 V_2 + \alpha_{27} C_u R_d P_m V_1 V_2$$

This multivariate linear model was fed in as hypothetical model for Regression Analysis in the SPSS software along with the parameter values from α_0 to α_{27} with their initial values set at "0". The mathematical model generated is as shown below:

The estimates of the coefficients α_0 to α_{27} can be found from the regression table C-2 in Appendix C. R squared = 0.88980

$$K_{star} = 8.412 + .179 C_u + 68.361 R_d + 3.461 P_m + -126.690 V_1 + 66.987 V_2 + 17.859 C_u R_d + 112.212 C_u P_m + 89.948 C_u V_1 - 62.486 C_u V_2 - 1954.010 R_d P_m - 486.586 R_d V_1 - 163.879 R_d V_2 - 410.497 P_m V_1 - 315.278 P_m V_2 + 60.811 V_1 V_2 + 1613.790 C_u R_d P_m + 348.459 C_u R_d V_1 + 84.907 C_u R_d V_2 + 109.411 C_u P_m V_1 - 268.367 C_u P_m V_2 + 13616.432 R_d P_m V_1 + 1984.610 R_d P_m V_2 + 344.927 R_d V_1 V_2 + 1552.438 P_m V_1 V_2 - 13137.139 C_u R_d P_m V_1 - 21755.670 R_d P_m V_1 V_2 + 15275.445 C_u R_d P_m V_1 V_2$$

$$K_{star} \geq 0$$

3.6 Performance Evaluation of the Models

After the development of the model we have to test it for its performance. If it predicts the dependent variable with sufficient accuracy it means the model is fit for use. We make use of past data or the data of which we know the value of all the factors and the dependent variable to test the model.

The optimization models can be tested if it gives us the better or optimal value of some performance evaluation measure. In the current study we have the reference model of EMSR. If the developed model gives better performance than EMSR, the model is fit to use.

Here two models are developed. A comparison has to be made between each model and EMSR model and also between each other. This can be done if we get the percentage number of times the developed model performed better than the existing model.

The % improvement data was collected using the simulation program. 500 random data points were provided to the simulation program. The simulation program was adjusted to give the average revenue value generated from the three models. A separate method was built for each model. Necessary information was passed to these methods. Each of this method then returned the protection level calculated from the developed model. These protection levels were used to calculate the average revenues, minimum revenues and maximum revenues. Table 3.7 shows a comparison of performance of EMSR model, Model1 and Model2.

Table 3.8 Comparison of performance of EMSR model Model1 and Model2

	Memsr	Model1	Model2
% no of times model is best	34.268	28.456	28.256
% loss in using the model w. r. t. EMSR	0	0.778545334	1.208896243
% advantage in using the model w. r. t. next best	0.31274069	0.03594376	0.02378671

Above results show that EMSR model is still performing better than the other two models. EMSR model gives best results about 34 % of the time whereas other two models are giving their best, approximately same no of times (28%). The percent loss of using the new models with respect to EMSR model is 0.78% for Model1 whereas it is higher for Model2 approximately 1.2 %. The percent advantage in using the model with respect to the next best is coming out to be 0.31 % for EMSR model Whereas the advantage is lower for other two models.

The results also suggest that the developed models are giving better results over EMSR model for certain number of times (though not all the time) and also can be advantageous if revised with some more facts. This implies that there are other factors which are not considered while developing the model but have profound effects on the average revenues. It thus suggests to study the problem more exhaustively and revise the developed models that will take into consideration the unknown parameters. Moreover regression analysis could be done with more data points.

So in this section we have studied the nature of average revenue with respect to changing protection levels. The three examples were studied each for different capacity utilization, which revealed the scope for revenue improvement. Design of experiments were performed considering first the μ_{f_1} / μ_{f_2} ratio and then the demand distributions of two fare classes. Significant factors were found out by Analysis of Variance and Regression Analysis was performed. In the process two models were developed and both these models were compared with EMSR model. At this initial stage the models performed well though not perfectly well and revealed that there is further scope or research to develop these models by considering still unknown factors.

Chapter 4

Conclusions and Future Work

4.1 Conclusion

Airline seat inventory management has been an attractive topic for research since the emergence of Revenue Management due to its potential to improve the revenue from each of the seat available in a segregated market using price discrimination. More attention has been paid to the optimization models used to allocate seats to different fare classes under very simple assumptions like no fare dispersion within a class rather than more realistic assumptions like wide dispersion of fares within classes.

The current study considered the fare distribution of the two classes so that the requests from potential low class customers paying higher fares should not be ignored. The simulated system developed during the study has explored the impact of these fare distributions on the average revenues computed by traditional EMSR method .. The behavior of average revenue was studied against different protection levels for a specific demand distribution and fare distribution which supported the idea to have lower protection levels for more dispersed fare distribution. Three scenarios with different capacity utilization were studied and found out that the revenue gains can be more that 0.5 % when capacity utilization is more than 1 and less than 0.5 % when capacity utilization is less than 1.

Design of Experiments and Analysis of variance were performed in two phases, first fully concentrating on the impact of fare distributions on average revenue and then considering factors associated with demand distributions. In the first phase μ_{f_1} / μ_{f_2} came out to be the most significant factor whereas in the second phase, capacity utilization, demand ratio, coefficients of variation of demand distributions along with overlapping area of two fare distributions (P_m), were the significant factors.

Regression Analysis was used to develop the models for optimal seat allocation. Two models were developed. First one computes the protection level considering five

cases of capacity utilization separately. The second model includes the capacity utilization as one of its variable and computes the protection level. A performance evaluation of the three models namely EMSR, Model1 and Model2 was done. It was observed that in 66% cases Model 1 or Model 2 gave better results than EMSR model. However individually models I or model II do not seems to give better result.

4.2 Future Work

The models can further be developed by considering other parameters affecting the optimal revenue which were not considered in this study. The study can further be extended to develop the models which will take into consideration more than two booking classes in a single leg seat environment. Moreover one can go still further by considering multiple leg scenario or entire network environment for a seat allocation problem to maximize revenues.

References

[1] P. P. Belobaba (1989): 'Application of a Probabilistic Decision Model to Airline Seat Inventory Control', Operations Research, 37, 2, 183–197.

[2] Larry Weatherford (2003): 'Dispersed Fares Within a Fare Class: How Can The Reality be Harnessed?', Journal of Revenue and Pricing Management, 3, 1, 26–40

[3] Larry Weatherford (2002): 'Simulated revenue impact of a new revenue management strategy under the assumption of realistic fare data', Journal of Revenue and Pricing Management, 1, 1, 35–49.

[4] J. I. McGill and G.J. Van Ryzin (1999): 'Revenue Management: Research Overview and Prospects', Transportation Science, 33, 2, 233–256.

[5] P. P. Belobaba (1992): 'Optimal vs. Heuristic Methods for Nested Seat Allocation', in Proceedings of AGIFORS Reservations and Yield Management Study Group, Brussels, Belgium, 28–53.

[6] Larry Weatherford (1998): 'A Tutorial on Optimization in the Context of Perishable-Asset Revenue Management Problems for the Airline Industry', Operations Research in the Airline Industry, Kluwer Academic Pub, 68–100.

[7] David Hassan (2003): 'The Need to Update Inventory Management Systems for Airlines—Part I', Journal of Revenue and Pricing Management, 2, 3, 234–244.

[8] David Hassan (2004): 'The Need to Update Inventory Management Systems for Airlines—Part II', Journal of Revenue and Pricing Management, 2, 4, 331–337.

[9] Sanne Vincent de Boer (2003): 'Advances in Airline Revenue Management and Pricing', Ph. D. Dissertation, MIT, Sloan School of Management.

[10] K. Littlewood (1972): 'Forecasting and Control of Passenger Bookings', in AGIFORS Symposium Proc. 12, Nathanya, Israel.

[11] M. J. Durham (1995): 'The Future of SABRE," in The Handbook of Airline Economics, D. Jenkins (ed.), The Aviation Weekly Group of the McGraw-Hill Companies, New York, NY, 485–491.

[12] K. T. Talluri and G. J. G.J. Van Ryzin (1999b); 'An Analysis of Bid-Price Controls for Network Revenue Management', *Management Sci.* 44, 1577–1593.

[13] W. M. Swan(1990): 'Revenue Management Forecasting Biases', Working Paper, Boeing Commercial Aircraft, Seattle, WA.

[14] A. O. Lee (1990): 'Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process', Ph.D. thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

[15] J. I. McGill (1995): 'Censored Regression Analysis of Multiclass Demand Data Subject to Joint Capacity Constraints', *Ann. Opns. Res.* 60, 209–240.

[16] J. D. Dana (1996): 'Peak-Load Pricing when the Peak Time is Unknown', Working Paper 96-84, General Motors Research Center for Strategy in Management, Kellogg School, Northwestern University, Evanston, IL.

[17] S. Borenstein and N. L. Rose(1994): 'Competition and Price Dispersion in the U.S. Airline Industry', *J. Political Economy* 102, 653–683.

[18] Larry Weatherford(1994): 'Optimization of Perishable-Asset Revenue Management Problems that Allow Prices as Decision Variables', Working Paper, University of Wyoming, Laramie, WY.

[19] G. Gallego (1996): 'A Demand Model for Yield Management', Technical Report, Columbia University, Dept. of Industrial Engineering and Operations Research, New York, NY

[20] R. E. Curry (1990): 'Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations', *Transp. Sci.* 24, 193–204.

[21] R. D. Wollmer (1986c): 'A Hub-Spoke Seat Management Model, unpublished company report', Douglas Aircraft Company, McDonnell Douglas Corporation, Long Beach, CA.

[22] B. C. Smith and C. W. Penn (1988): 'Analysis of Alternative Origin–Destination Control Strategies', in *AGIFORS Symposium Proc.* 28, New Seabury, MA.

[23] R. W. Simpson (1989): 'Using Network Flow Techniques to Find Shadow Prices for Market and Seat Inventory Control', Memorandum M89-1, MIT Flight Transportation Laboratory, Cambridge, MA.

[24] E. L. Williamson(1992): 'Airline Network Seat Inventory Control: Methodologies and Revenue Impacts', Ph.D. thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

[25] P. P. Belobaba (1987a): 'Air Travel Demand and Airline Seat Inventory Management' Ph.D. thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

[26] www. indian-airlines.nic.in

[27] www.jetairways.com

[28] www.airsahara.net

[29] www.airdeccan.net

[30] www.flykingfisher.com

[31] www.spicejet.com

Appendix A

A brief note on fare classes in Airline Industry

Sometimes people get confused with the travel class and the fare class. This appendix is just to give a clear cut idea of Airline codes for fare classes.

Airlines

Airlines traditionally have three travel classes:

- First class, the highest quality of accommodation available
- Business class, high quality, traditionally purchased by business travellers
- Economy class (also known as coach or tourist class), basic accommodation, commonly purchased by leisure travellers

However, with premium travelers choosing private planes and businesses becoming more fare-sensitive, the classes have become blurred. Three-cabin aircraft are found only on premium transcontinental or international flights. For shorter distances, most airlines fly a two-cabin plane, featuring a downscaled first class which may be comparable to international business, to accompany economy. Some airlines merge their international first and business classes into a premium business product, whereas others either supplement or supplant the business cabin with a premium economy class. Some exceptionally long flights operated by Singapore Airlines offer only Business and Premium Economy class service.

But most notable, most low-cost carriers have eliminated their first and business class cabins entirely. The costs of extra services and amenities afforded to the premium cabins is eliminated, and more seats can be installed on an aircraft.

Fare class

Within each travel class there are often different *fare classes*, relating to ticket or reservation restrictions and used to enhance opportunities for price discrimination. Passengers within the same travel class receive the same quality of accommodation and may indeed sit next to each other; however, the price or restrictions they face for that accommodation will vary depending on the fare class. For example, full fare economy class passengers (booking code Y) are usually able to make changes to their

reservation, while discount economy class passengers in the lowest booking code usually have tickets that are non-refundable, non-upgradeable, non-transferable, or non-changeable without a hefty fee.

Airline fare classes are commonly indicated by letter codes, but the exact hierarchy and terms of these booking codes vary greatly from carrier to carrier.

First class codes

- A, F, P, R, Z

On domestic flights, F commonly indicates first class on a two-cabin plane. If a three-cabin aircraft is used, P (for "premium") may be used to distinguish the higher level of service in first class. The R code indicated supersonic transport and is no longer used with the retirement of the Concorde. The A and Z codes may indicate a first class ticket whose fare is reduced due to restrictions on refunds, advance reservation requirements, or other terms.

Business class codes

- C, J, D, I

On many airlines, C or J indicate full fare business class, whereas discounted and thus restricted and un-upgradeable fares are represented by D or I.

Economy class codes

- Full fare: Y, H
- Standard fare: B, M, N
- Special or discount fares: G, I, K, L, O, Q, S, T, U, V, W, X

In most airlines, a full fare, unrestricted economy ticket is booked as a Y fare. Full fare tickets with restrictions on travel dates, refunds, or advance reservations are commonly classed as B, H, or M, although some airlines may use S, W, or others. Heavily discounted fares, commonly T or W, not only will not permit cabin upgrades, refunds, or reservation changes, but may require Internet booking, not credit miles or elite status credit in a frequent flyer program, or impose many other restrictions. Other fare codes are restricted for use by consolidators, group charters, or travel industry professionals.

Most low-cost carriers have greatly simplified the fare classes they use to a handful of cases, unlike the dozens employed by a traditional airline. While some traditional carriers have followed, others continue to prefer price discrimination over commoditization.

Source : http://en.wikipedia.org

The fare structure of Jet Airways looks like this:

Club Premiere	J
Economy	L Q S K H

The fare structure of Indian Airlines looks like this:

First Class:	A, F
Business Class:	C, D, J
Economy Class:	H, K, L, M, O, N, S, V, Q

The fare structure of Air Sahara looks like this:

Business Class:	J, D
Economy Class:	Y, K, U, T, E, M, N, O, Q

The Air Deccan and Kingfisher Airlines have only one travel class with more than one booking classes

Appendix B

Revenue Management Glossary

J. I. McGill and G.J. Van Ryzin (1999) [4]

We Provide here a glossary of the sometimes-confusing terminology of revenue management. Our aim in supplying a separate glossary is to avoid needless definitions for readers familiar with revenue management while assisting others who are new to the field. Many of the terms described here have different meanings in more general contexts but are presented here with their usual meanings in revenue management.

Aggregation of demand: The level of summarization of passenger demand data. The trend has been toward increasing levels of disaggregation in seat inventory optimization; however, pricing, forecasting, and booking control processes often operate at different levels of aggregation. Possible dimensions for disaggregation include: market, season, month, week, section of week (e.g., midweek versus weekend), day of week, time of day, *flight number*, booking class, fare, *flight leg*, segment, and itinerary.

Arrival pattern: The pattern of arrivals of booking requests. In the airline context, some possible arrival patterns are: sequential booking classes, low-before-high fares, or *interspersed arrivals*.

Batch booking: (also multiple booking, or bulk arrival) A booking request that arrives through normal reservation channels for two or more seats to be booked for the same itinerary. Contrast with *groupbookings*.

Bid price: A net value (bid-price) for an *incremental seat* on a particular flight leg in the airline network. Also referred to as *minimum acceptable fare*, hurdle price, probabilistic shadow price, displacement cost, or probabilistic dual cost.

Bid price control: A method of network seat inventory control that assesses the value of an ODF itinerary as the sum of the bid-prices assigned to individual legs in the itinerary. Typically, an ODF request is accepted if its fare exceeds the total bidprices. Also referred to as *continuous nesting*.

Booking class: A category of bookings that share common features (e.g., similar revenue values or restrictions) and are controlled as one class. This term is often used interchangeably with *fare class* or *bucket.*

Booking limit: The maximum number of seats that can be sold to a particular booking class. In *nested booking* systems, booking limits apply to the total number of seats sold to a particular booking class and any lower fare booking classes.

Booking policy: A booking policy is a set of rules that specify at any point during the booking process whether a booking class should be open. In general, such policies may depend on the pattern of prior demands or be randomized in some manner and must be generated dynamically as the booking process unfolds for each flight. In some circumstances, optimal or approximately optimal booking policies can be defined by a set of fixed *protection levels* or *threshold curves.*

Buckets: This term is used in two related ways. First, in older reservations systems, seats for different fare classes or groups of classes are pre-assigned to distinct buckets. These seats are available exclusively to bookings in that fare class. This method simplifies reservations control but is clearly undesirable from a revenue standpoint because seats could fly empty in a discount bucket even if there is higher fare demand available to fill them. Second, buckets also refer to clusters of different fare classes or ODFs that are grouped together for control purposes in a *virtual nesting* system. A single booking limit is set for all classes in the bucket or lower value buckets.

Bulk arrival: See *batch booking.*

***Bumping:** See denied boarding.*

Cabin: The physical compartment of an aircraft containing a particular type of seating. For example, an aircraft may be equipped with a first class cabin and a coach cabin, each with different seating and separated by a partition. Multiple fare classes are usually available in each cabin of the aircraft.

Cancellations: Returns or changes in bookings that occur early enough in the booking period to permit subsequent rebooking through the reservations system.

Censorship of demand data: Typically, no record can be kept of booking requests that occur after a fare class is closed down. Thus, in booking histories, the number of flights on which demand reached a booking limit can be determined but not the

amount by which demand exceeded the limit. Formally, this condition is known as multiple Type I censorship of the data—the censorship points are known (booking limits), but may vary between observations (flights). Code-sharing: It is now relatively common for small groups of domestic and international airlines to form alliances in which the members interlist some or all of their flights. The two character airline designator code from one airline is applied to flight numbers of other alliance airlines so that there is an apparent expansion of participating airlines' networks.

Coefficient of variation: The standard deviation expressed as a proportion of the mean of a probability or relative frequency distribution. Thus a *demand distribution* with mean demand 100 and standard deviation 40 would exhibit a coefficient of variation of 0.40. Airline demand data typically display coefficients of variation in the range 0.25 to over 1.0, depending on the level of aggregation of the data.

Connectivity (in reservations systems): The degree to which the elements of the reservations system are electronically interconnected. See *seamless availability*.

Continuous nesting: (see *bid-price control*)

Controllable booking classes: All early reservations systems and many existing systems offer only a small number of distinct booking categories (five to ten) that can actually be controlled at booking outlets. Thus, regardless of the number of booking classes or distinct passenger itineraries that can be handled by the revenue management optimization process, the controls in such systems can only be applied to a small number of aggregate booking classes or buckets.

Control limit policy: A structural solution that specifies an upper bound (limit) on the number of seats sold in each fare class (or collection of fare classes) for each time before flight departure.

CRS: Computer reservations system.

Defections: It can occur that a confirmed passenger who shows up for a flight switches to a flight with another airline (usually because of a delay in the original flight departure). Defections constitute a relatively small component of lost passengers and are normally counted as part of no-shows. However, they are distinct from no-shows, and any attempt to predict their occurrence requires an estimation of the probability distribution for departure delays.

Demand distribution: An assignment of probabilities (probability distribution) to each possible level of demand for a flight or booking class. A preliminary estimate of such a demand distribution can be obtained by calculating the proportion of each demand level seen on comparable past flights; i.e., a relative frequency distribution.

Demand factor: The ratio of demand over capacity for a flight or booking class. (Contrast with *loadfactor.)*

Denied boarding: Turning away ticketed passengers when more passengers show-up at flight time than there are seats available on the flight, usually as a result of overbooking practices. Denied boardings can be either **voluntary**, when passengers accept compensation for waiting for a later flight, or **involuntary**, when an insufficient number of passengers agree to accept compensation. In the latter case, the airline will be required to provide compensation in a form mandated by civil aviation law.

Disaggregate: *See aggregation of demand.*

Displacement cost: In revenue management, the displacement (or *opportunity*) cost of a booking includes all future revenues that may be lost if the booking is accepted. Taken to the extreme, these include the revenue value of potential displaced future bookings anywhere in the airline network and *goodwill costs* from those displacements. Assessment of the costs and probabilities of such displacements should allow for the dynamics of cancellations and overbooking and the expected costs of oversold conditions.

Diversion: The booking of a customer at a fare level lower than one they would have been prepared to pay. This occurs, for example, when a business traveler has sufficient advance notice of a trip to book in a discount class intended primarily for leisure travelers. *Restrictions* are designed to inhibit such diversion.

Dual prices (also shadow prices): The marginal value of one additional unit of a constrained resource, as determined by a mathematical programming solution to an optimization model. Dual prices are one source of the marginal seat values used in bid-price control.

Dynamic models: Models that take into account future possible booking decisions in assessing current decisions. Most revenue management problems are properly modeled as dynamic programming problems.

Expected marginal seat revenue (EMSR): The expected revenue of an incremental seat if held open. This is a similar concept to that of bid-price but generally used in a simpler context. Expected revenue: The statistical expected revenue; that is, the sum of possible revenue values weighted by their probabilities of occurrence. Fare basis code: An alphanumeric encryption of the conditions and restrictions associated with a given fare. Usually several fare basis codes are contained in a single fare class.

Fare class: A category of booking with a (relatively) common fare. Typical labels for such classes [see Vinod (1995)] are: **F** for first class (separate compartment);

J for business class, **U** for business class frequent flyer redemption (often separate compartment); **Y** for full fare coach; **B, M, Q, V** for progressively more discounted coach bookings; and **T** for frequent flyer coach cabin redemptions. Often other fare products (such as travel agent or company travelers) are categorized under one of these designations for control purposes.

Fare product: The full set of attributes associated with a specific transportation service. The set includes the fare as well as any restrictions or benefits that apply to that service at that fare.

Fences: *See restrictions.*

Fleet assignment: Most airlines have a variety of aircraft types and sizes in their fleets. The fleet assignment process attempts to allocate aircraft to routes in the airline network to maximize contribution to profit. There are strong potential linkages between fleet assignment and revenue management processes because aircraft assignments determine leg capacities in the network.

Flight leg: A section of a flight involving a single takeoff and landing (or no boarding or deplaning of passengers at any intermediate stops). Also *leg*.

Flight number: A numeric or alphanumeric label for a flight service that involves (generally) a single aircraft departing from an origin airport, possibly making additional scheduled stops at one or more intermediate airports, and terminating at a destination airport.

Full Nesting: See *nested booking.*

Global distribution system (GDS): Computer and communications systems for linking booking locations with the computer reservation systems of different airlines. Examples are SABRE, Galileo, and Amadeus.

Goodwill costs: An airline's rejection of a booking request can affect a customer's propensity to seek future bookings from that airline. This cost is difficult to assess but is considered particularly acute in competitive markets and with customers who are frequent air travelers. An approximate assessment of the cost of a permanently lost customer is the expected net present value of all future bookings from the customer minus the opportunity costs of those bookings.

Go-show: Passengers who appear at the time of flight departure with a valid ticket for the flight but for whom there is no record in the reservation system. This no-record situation can occur when there are significant time lags in transferring booking information from reservations sources (e.g., travel agent's offices) to the CRS or when there are transmission breakdowns.

Group bookings: Bookings for groups of passengers that are negotiated with sales representatives of airlines; for example, for a large group from one company travelling to a trade show. These should be distinguished from batch bookings.

Hub-and-spoke network: A configuration of an airline's network around one or more major hubs that serve as switching points in passengers' itineraries to spokes connected to smaller centers. The proliferation of these networks has greatly increased the number of passenger itineraries that include connections to different flights.

Hub bank: A collection of inbound and outbound flights that are scheduled to arrive or depart within a time span that enables convenient passenger connections among flights. An airline hub will typically operate with several hub banks throughout the day. **Incremental seat:** One additional seat, given the number of seats already booked.

Independence of demands: The assumption that demands in one customer category (e.g., booking class or ODF) are statistically independent of demands in other categories. It is widely believed that this assumption is not satisfied in practice. See, for example, Hopperstad (1994).

Indexing: The process of assigning individual ODF categories to virtual nesting buckets. Smith, Leimkuhler, and Darrow (1992) provide details.

Interspersed arrivals: Characteristic of an arrivals process in which booking requests in different booking classes do not arrive in any particular order. (Compare with sequential booking classes.)

Itinerary: For purposes of this paper, an itinerary is a trip from an origin to a destination across one or more airline networks. A complete specification of an itinerary includes departure and arrival times, flight numbers, and booking classes. The term is used ambiguously to include both one-way and round-trip travel. That is, used in the first way, a round-trip involves two itineraries and, in the second way, one itinerary.

Leg: *See flight leg.*

Leg based control: An older, but still common, method of reservations control and revenue management in which limits are set at the flight leg level on the number of passengers flying in each booking class. Such systems are unable to properly control *multileg* traffic, although virtual nesting provides a partial solution.

Littlewood's rule: This simple two-fare allocation rule was proposed by Littlewood (1972). Given average high fare *f1*, average discount fare *f2*, random full fare demand *Y*, and *s* seats remaining, Littlewood's rule stipulates that a discount seat should be sold as long as the discount fare equals or exceeds the expected marginal return from a full fare booking of the last remaining seat; that is, discount demand should be satisfied as long as *f2* $ *f1* Pr(*Y* . *s*). This is essentially equivalent to the classic optimal stocking rule for single period stochastic inventory (newsvendor) problems.

Load factor: The ratio of seats filled on a flight to the total number of seats available.

Low-before-high fares: (Also called *monotonic fares* or *sequential fares.*) The sequential booking class assumption is often augmented by the additional assumption that booking requests arrive in strict fare sequence, generally from lowest to highest as flight departure approaches. The existence of low *standby fares* violates this assumption.

Minimum acceptable fare (MAF): See *bid-price.*

Monotonic fares: *See low-before-high fares.*

Multileg: A section of an itinerary or network involving more than one *leg.*

Multiple booking: See *batch booking.*

Nested booking: In fully nested (also called *serially nested*) booking systems, seats that are available for sale to a particular booking class are also available to bookings in any higher fare booking class, but not the reverse. Thus, a *booking limit* L for a discount booking class defines an upper bound on bookings in that class and any lower valued classes and a corresponding protection level of (C 2 L) for all higher classes; where C is the total capacity of the pool of seats shared by all classes. This should be contrasted with the older distinct bucket approach to booking control. See, also, *parallel nesting.*

Network effects: A booking on any leg in the airline network may block booking of any itinerary that includes that leg. Subsequent interactions of the blocked itinerary with other legs in the network can, in a similar fashion, propagate across the full network.

Newsvendor problem: The problem of choosing the quantity of a perishable item to stock (e.g., newspapers) given known cost, selling price, and salvage values, and subject to uncertain future demand. (Also called the newsboy or single period stocking problem.) This classic problem is essentially equivalent to the simple two-fare *seat allocation problem* with sequential arrivals.

No-shows: Booked passengers who fail to show up at the time of flight departure, thus allowing no time for their seat to be booked through normal reservations processes. No-shows are particularly common among full fare passengers whose tickets are fully refundable in the event of cancellation or no-show.

***OBL:** See optimal booking limits.*

ODF control (O–D problem): Origin–destination fare control. An approach to revenue management that accounts for all possible passenger itineraries between origins and destinations in the airline network, at all fare levels. See *network effects.* Opportunity cost: See *displacement cost.* Optimal booking limits: This term is often used to refer to exact booking limits for the single leg seat inventory control under assumptions 1 through 6 in Section 4.1. They are only optimal within the context of that basic model. At present, there are no truly optimal booking limits for the full ODF revenue management problem, and likely never will be.

Origin–destination control: See *ODF control.*

Overbooking: The practice of ticketing seats beyond the capacity of an aircraft to allow for the probability of no-shows.

Oversold: An ambiguous term sometimes used when more passengers show up for a flight than there are seats available. Such situations must be resolved with denied boardings.

Parallel nesting: See *nested booking.* This is an approach to booking that is intermediate between simple distinct bucket control and *full nesting.* A number of lower fare classes are assigned to distinct buckets, but these buckets are nested in one or more higher fare classes. This approach reduces the revenue potential of the combined fare classes, but may facilitate control.

Perishable asset revenue management (PARM): A term introduced in Weatherford and Bodily (1992) for the general class of revenue management problems, which includes airline revenue management.

Protected seats: Seats that are restricted to bookings in one or more fare classes. In fully nested booking systems, seats are protected for bookings in a fare class or any higher fare class.

Protection levels: The total number of protected seats for a booking class. In fully nested booking systems the protection level for a fare class applies to that class and all higher fare classes.

RCS: Reservations Control System.

Recapture: The booking of a passenger who is unable to obtain a reservation for a particular flight or 249 RESEARCH OVERVIEW AND PROSPECTS / set of flights with an airline onto alternative flights with the same airline.

Reservation system controls: The internal logic used by the reservation system for controlling the availability of seats. This logic is usually difficult to change and is often a significant constraint when implementing a yield management system. See *controllable booking classes.*

Restrictions: Sets of requirements that are applied to discount fare classes to differentiate them as fare products and discourage diversion. Examples are fourteen-day advance booking requirements, cancellation penalties, Saturday night stayover, and midweek departure requirements. Also referred to as *booking fences.*

Revenue management: The practice of controlling the availability and/or pricing of travel seats in different booking classes with the goal of maximizing expected revenues or profits. This term has largely replaced the original term yield management.

Rules: *See restrictions.*

Seamless availability: A capability of reservation and information systems that allows for direct transmission of availability requests from ticket agents to airlines. With this capability, airlines may be able to provide unrestricted origin– destination control of their seat inventory.

Seat allocation: See *seat inventory control.*

Seat inventory control: The component of a revenue management system that controls the availability of seats for different booking classes.

Segment: One or more flight legs covered by a single flight number. Thus, if a flight originates at airport **A**, makes an intermediate stop at **B**, and terminates at **C**; the possible flight segments are **AB**, **BC**, and **ABC**.

Segment closed indicator (SCI): A flag in reservations control systems that indicates that a booking class is closed to bookings over a particular segment. The same booking class may be open for bookings over other segments of the same flight. This allows for O–D control at the segment level.

Segment control: A level of itinerary seat inventory control that accounts for the revenue value of flight segments, but does not account for itineraries that involve other flight segments. In the case of a two leg flight **A** to **B** to **C**, segment control would permit closing the **AB** segment to a discount booking class but leaving the **ABC** segment open for the same class. This system fails to account for the possibly high revenue value of a booking that includes, for example, the segment **AB** in its itinerary but switches to a different flight at **B**.

Sequential booking classes: The assumption that requests for bookings in particular classes are not interleaved; for example, all B-class requests arrive before any Y-class requests. This assumption is rarely satisfied in practice; however, it is close enough to permit significant revenue gains from methods based on the assumption. Also, early booking restrictions on many discount booking classes ensure a degree of compliance.

Sequential fares: *See low-before-high fares.*

Serial nesting: See *nested booking.*

Show-ups: Passengers who appear for boarding at the time of flight departure. The number of showups is (final bookings 1 go-shows 1 standbys 2 no-shows).

Single-leg control: See *leg based control.*

Space control: See seat inventory control.

Spill: Unsatisfied demand that occurs because a

capacity or booking limit has been reached. See *censorship*

of demand data.

Spill formula: A formula or algorithm that estimates the amount of spill that has occurred on past flights.

Spoilage: Seats that travel empty despite the presence of sufficient demand to fill them. This will occur, for example, if discount booking classes are closed too early,

and full fare demands do not fill the remaining seats. This should be distinguished from excess capacity—seats that are empty because of insufficient total demand.

Standby fares: Some airlines will sell last minute discount seats to certain categories of travelers (e.g., youth or military service personnel) who are willing to wait for a flight that would otherwise depart with empty seats.

Static models: Models that set current seat protection policies without consideration of the possibility of adjustments to the protection levels later in the booking process. (Compare with *dynamic models.*)

Structural solution: A solution to an optimization problem in the form of specifications (frequently equations) that reveal the pattern of behavior of optimal solutions. These are important because they lead to a deeper understanding of the nature of optimal solutions and can lead to development of efficient solution algorithms.

Threshold curves: Threshold curves are functions that return time-dependent booking limits for overbooking or seat inventory control.

Unconstrained demand: An estimate of the demand for a past flight or fare class that has been corrected for censorship.

Upgrade: This term is used in two ways. First, it refers to an offer to a passenger to fly in a higher service class without additional charge (e.g., in exchange for frequent flyer points, or to avoid a denied boarding). Second, it refers to a decision by a customer to book in a higher fare class than originally intended when he or she is advised that no seats are available at their preferred fare.

Virtual nesting/virtual classes: This is one approach to incorporating origin–destination information into leg or segment based control systems. Multiple ODFs are grouped into virtual buckets on the basis of similar revenue characteristics (e.g., comparable total fare values, or similar total bid prices). The virtual buckets may easily contain a mixture of traditional fare classes. The buckets are then nested and assigned to traditional booking classes for control in a leg based reservation system.

Yield management: The early term used for what is now more commonly called revenue management. Cross (1995) attributes the original term to Robert L. Crandall when he was Senior Vice President for Marketing (later CEO) at American Airlines.

Weatherford and Bodily (1992) introduced the general term, perishable asset revenue management, for the general class of inventory control problems of which airline revenue management is an example.

Appendix C

Regression tables produced by SPSS software

Table C - 0.8

Regression table for Model1 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	16	18755.61010	1172.22563
Residual	84	299.58990	3.56655
Uncorrected Total	100	19055.20000	
(Corrected Total)	99	2916.03840	

R squared = 1 - Residual SS / Corrected SS = .89726

			Asymptotic 95 % Confidence Interval	
Parameter	Estimate	Asymptotic Std. Error	Lower	Upper
α_0	32.369451112	42.458489082	-52.06391072	116.80281295
α_1	-63.94252476	100.10977820	-263.0217974	135.13674793
α_2	-413.9278569	493.45916505	-1395.225523	567.36980902
α_3	-158.6018795	156.48454746	-469.7885633	152.58480434
α_4	-111.0240817	146.91540866	-403.1814836	181.13332019
α_5	1215.4183423	1421.6534268	-1611.695405	4042.5320899
α_6	465.87399534	367.73282816	-265.4030616	1197.1510523
α_7	464.07470430	360.57678479	-252.9717756	1181.1211842
α_8	1550.6448435	1846.7687813	-2121.857408	5223.1470948
α_9	1769.6748371	1833.6094447	-1876.658630	5416.0083042
α_{10}	591.12829713	545.44906712	-493.5569906	1675.8135848
α_{11}	-4208.913721	5438.8691315	-15024.70145	6606.8740118
α_{12}	-5633.053712	5371.7510785	-16315.36984	5049.2624109
α_{13}	-2134.737650	1336.5637455	-4792.641235	523.16593488
α_{14}	-6752.903873	6927.9227990	-20529.83812	7024.0303772
α_{15}	19459.733178	20716.797654	-21737.89091	60657.357262

Table C - 0.9

Regression table for Model1 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	16	25454.00731	1590.87546
Residual	84	815.88269	9.71289
Uncorrected Total	100	26269.89000	
(Corrected Total)	99	5282.57310	

R squared = 1 - Residual SS / Corrected SS = .84555

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
α_0	79.668686115	54.708832561	-29.12582719	188.46319942
α_1	-239.8312148	158.52466745	-555.0749009	75.412471361
α_2	-92.64017795	679.77955176	-1444.456366	1259.1760101
α_3	-308.8608312	210.61066743	-727.6832404	109.96157805
α_4	-261.9094124	204.74020179	-669.0577569	145.23893213
α_5	661.89178153	2004.4825591	-3324.241613	4648.0251759
α_6	1013.4487255	622.53875482	-224.5378604	2251.4353114
α_7	1175.4099335	602.55717027	-22.84108017	2373.6609472
α_8	62.573793058	2800.9119445	-5507.346776	5632.4943623
α_9	477.26801106	2479.0312689	-4452.557535	5407.0935572
α_{10}	1065.2528757	798.84003788	-523.3281460	2653.8338973
α_{11}	-1171.706099	8276.4327946	-17630.30036	15286.888164
α_{12}	-3875.597073	7461.3979153	-18713.40510	10962.210951
α_{13}	-4285.613209	2388.1706976	-9034.752545	463.52612642
α_{14}	-736.0295147	10375.248059	-21368.34810	19896.289074
α_{15}	8199.0356543	31059.736670	-53566.65694	69964.728251

Table C - 1.0

Regression table for Model1 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	16	44167.76427	2760.48527
Residual	84	1754.93573	20.89209
Uncorrected Total	100	45922.70000	
(Corrected Total)	99	11095.67560	

R squared = 1 - Residual SS / Corrected SS = .84184

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
α_0	-61.02518314	90.753512022	-241.4984945	119.44812818
α_1	400.43972115	230.61504831	-58.16359326	859.04303556
α_2	-47.67028136	1219.6880911	-2473.153810	2377.8132473
α_3	256.96906818	337.76652257	-414.7167038	928.65484013
α_4	207.47561446	324.72758824	-438.2808067	853.23203558
α_5	192.86629576	3161.9549872	-6095.027958	6480.7605499
α_6	-1454.846745	862.07809133	-3169.183571	259.49008081
α_7	-973.8209145	833.55799347	-2631.442398	683.80056929
α_8	30.622358461	4057.1860430	-8037.537027	8098.7817444
α_9	739.02411164	4656.9590294	-8521.849633	9999.8978562
α_{10}	-807.8305581	1216.3926195	-3226.760680	1611.0995640
α_{11}	-344.4881861	10897.645054	-22015.65049	21326.674116
α_{12}	-2591.690102	12100.853274	-26655.56390	21472.183697
α_{13}	4169.0136720	3126.1942733	-2047.766481	10385.793825
α_{14}	-2517.048429	15568.500206	-33476.71844	28442.621581
α_{15}	6612.4369335	41883.896429	-76678.28440	89903.158262

Table C - 1.1

Regression table for Model1 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	16	70452.37358	4403.27335
Residual	84	793.46642	9.44603
Uncorrected Total	100	71245.84000	
(Corrected Total)	99	14143.95840	

R squared = 1 - Residual SS / Corrected SS = .94390

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
α_0	-15.10101705	49.921930917	-114.3762515	84.174217368
α_1	144.15341454	148.85474892	-151.8605781	440.16740722
α_2	783.70368089	562.55485419	-334.9983404	1902.4057021
α_3	80.340452772	192.29825770	-302.0657214	462.74662698
α_4	95.276589453	183.36073782	-269.3563463	459.90952523
α_5	-3663.805082	1577.8111466	-6801.455581	-526.1545835
α_6	-363.6676009	585.44155720	-1527.882341	800.54713918
α_7	-351.0465776	555.29563143	-1455.312838	753.21968305
α_8	-3029.747414	2073.9493640	-7154.023168	1094.5283398
α_9	-3742.979090	2095.8489107	-7910.804495	424.84631415
α_{10}	-382.2899889	704.72370546	-1783.710362	1019.1303843
α_{11}	13618.585823	5972.8736076	1740.8716271	25496.300018
α_{12}	17053.957415	6032.3527195	5057.9624824	29049.952348
α_{13}	1644.8475136	2178.5667906	-2687.471466	5977.1664936
α_{14}	14537.962860	7699.4097534	-773.1578052	29849.083526
α_{15}	-67798.83238	22999.433656	-113535.7285	-22061.93628

Table C - 1.2

Regression table for Model1 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	16	84858.37923	5303.64870
Residual	84	510.32077	6.07525
Uncorrected Total	100	85368.70000	
(Corrected Total)	99	15767.70760	

R squared = 1 - Residual SS / Corrected SS = .96764

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
α_0	-41.40372622	61.708449979	-164.1177464	81.310293946
α_1	224.34632494	167.60274501	-108.9501140	557.64276387
α_2	886.66779236	731.49860546	-567.9974058	2341.3329906
α_3	156.69621678	225.52493990	-291.7848588	605.17729240
α_4	160.08625446	223.42826656	-284.2253563	604.39786523
α_5	-393.7006556	2061.9597956	-4494.133838	3706.7325267
α_6	-530.3287372	624.78657161	-1772.785353	712.12787894
α_7	-514.3448022	599.30187050	-1706.122295	677.43269094
α_8	-3053.097424	2740.4736119	-8502.829741	2396.6348923
α_9	-3105.923298	2595.4986700	-8267.357044	2055.5104479
α_{10}	-520.1011742	818.37149189	-2147.522634	1107.3202858
α_{11}	-1060.463842	7836.9636771	-16645.12557	14524.197886
α_{12}	2928.7024074	7237.2735816	-11463.40980	17320.814614
α_{13}	1861.7666439	2231.1418638	-2575.103635	6298.6369226
α_{14}	10516.763737	9758.3763602	-8888.837827	29922.365301
α_{15}	-5268.755774	27638.956852	-60231.85256	49694.341007

Table C - 2

Regression table for Model2 produced by SPSS software

Source	DF	Sum of Squares	Mean Square
Regression	28	463795.71586	16564.13271
Residual	972	13314.94414	13.69850
Uncorrected Total	1000	477110.66000	
(Corrected Total)	999	120822.38464	

R squared = 1 - Residual SS / Corrected SS = .88980

Parameter	Estimate	Asymptotic Std. Error	Asymptotic 95 % Confidence Interval Lower	Upper
α_1	.179017523	33.707823219	-65.96947037	66.327505416
α_2	68.361305304	109.60197529	-146.7224421	283.44505273
α_3	3.461110475	339.92064425	-663.6017394	670.52396038
α_4	-126.6900072	125.72906835	-373.4216844	120.04166990
α_5	66.987310391	123.50079092	-175.3715782	309.34619900
α_6	17.859507378	90.973019143	-160.6666352	196.38565000
α_7	112.21278711	282.88971036	-442.9321257	667.35769988
α_8	89.948563790	92.437003874	-91.45051357	271.34764115
α_9	-62.48638982	86.464097018	-232.1641895	107.19140983
α_{10}	-1954.010055	802.45832813	-3528.760363	-379.2597468
α_{11}	-486.5865294	323.04423236	-1120.530981	147.35792187
α_{12}	-163.8790637	325.99678612	-803.6176288	475.85950145
α_{13}	-410.4978155	1082.3214389	-2534.453614	1713.4579835
α_{14}	-315.2784178	1080.2032453	-2435.077458	1804.5206220
α_{15}	60.811054624	314.91912606	-557.1886264	678.81073565
α_{16}	1613.7905602	566.27418684	502.52980182	2725.0513185
α_{17}	348.45928062	238.68968098	-119.9471593	816.86572057
α_{18}	84.907035676	230.60775555	-367.6393725	537.45344387
α_{19}	109.41102504	801.84030878	-1464.126477	1682.9485274
α_{20}	-268.3671281	794.57262393	-1827.642471	1290.9082144
α_{21}	13616.432087	3575.0259258	6600.7741004	20632.090073
α_{22}	1984.6107395	2066.1099817	-2069.939156	6039.1606350
α_{23}	344.92795094	812.98379093	-1250.477605	1940.3335071
α_{24}	1552.4384501	2760.8732188	-3865.520090	6970.3969905
α_{25}	-13137.13961	2999.9846054	-19024.33215	-7249.947076

α_{26}	-21755.67034	10646.146008	-42647.74799	-863.5926827
α_{27}	15275.445210	7892.0311796	-211.9366223	30762.827043
α_{0}	8.412065570	41.146813375	-72.33475302	89.158884156